

---

Electronic Thesis and Dissertation Repository

---

9-13-2013 12:00 AM

# The Process Dissociation of Moral Judgments: Clarifying the Psychology of Deontology and Utilitarianism

Paul Conway

*The University of Western Ontario*

Supervisor

Dr. Jim Olson

*The University of Western Ontario* Joint Supervisor

Dr. Bertram Gawronski

*The University of Western Ontario*

Graduate Program in Psychology

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Paul Conway 2013

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Social Psychology Commons](#)

---

## Recommended Citation

Conway, Paul, "The Process Dissociation of Moral Judgments: Clarifying the Psychology of Deontology and Utilitarianism" (2013). *Electronic Thesis and Dissertation Repository*. 1630.

<https://ir.lib.uwo.ca/etd/1630>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

THE PROCESS DISSOCIATION OF MORAL JUDGMENTS: CLARIFYING THE  
PSYCHOLOGY OF DEONTOLOGY AND UTILITARIANISM

(Thesis format: Integrated Article)

by

Paul Conway

Graduate Program in Psychology

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy in Psychology

The School of Graduate and Postdoctoral Studies  
The University of Western Ontario  
London, Ontario, Canada

© Paul Conway 2013

## Abstract

A growing body of work has examined responses to moral dilemmas where causing some degree of harm leads to a greater positive outcome; such dilemmas are said to pit *deontological* philosophical considerations (causing harm is never acceptable) against *utilitarian* philosophical considerations (causing harm is acceptable if it leads to the best possible outcome). According to dual-process theories of moral judgment, independent processes drive each judgment: affective reactions to harm drive deontological judgments, whereas cognitive evaluations of outcomes drive utilitarian judgments. Yet, theoretically both processes contribute to each judgment; therefore, it is an error to equate judgments with processes. To overcome this error, we adapted Jacoby's (1991) *process dissociation* (PD) procedure to independently quantify the strength of deontological and utilitarian inclinations within individuals. Five studies presented in two articles support the conclusion that process dissociation taps the processes theorized to underlie moral judgments more effectively than overt dilemma judgments, and allows for increased insight into the nature of moral processing. In Conway and Gawronski (2013) Study 1, the PD parameters predicted theoretically relevant individual-difference variables (e.g., the utilitarian parameter uniquely predicted cognitive load, whereas the deontology parameter uniquely predicted empathic concern and perspective-taking). Moreover, both parameters predicted moral identity—a relation that was obscured using overt moral judgments. In Study 2, a cognitive load manipulation selectively reduced utilitarian inclinations, whereas in Study 3, a manipulation that increased the vividness of harm selectively increased the deontology parameter. Together, these findings suggest that the deontology parameter is tapping affective reactions to harm, and the utilitarian parameter is tapping cognitive evaluations of outcomes (consistent with theory). In Study 1 of Conway, Bartels, and Pizarro (under review), participants scoring higher in Machiavellianism, psychopathy, and meaninglessness made more overt utilitarian judgments (replicating past findings), but process dissociation revealed that this relation was due to *decreased deontology* rather than *increased utilitarianism* among people high in antisocial personality traits. Study 2 demonstrated that the deontology and utilitarian

parameters each correlated with different kinds of prosociality. These findings clarify theoretical confusion regarding the nature of utilitarianism and deontology.

## Keywords

Morality, judgments and decision-making, process dissociation, emotion, affect, cognition, psychopathy, prosocial behavior, moral dilemmas, prosociality, ethics,

## Co-Authorship Statement

Several co-authors contributed to the work contained in this dissertation. The first article was co-authored by Paul Conway and Bertram Gawronski. Both authors contributed to the generation of the research questions, the design of experiments, the interpretation of results, and the publication of those results. However, Paul Conway was first author and was primarily responsible for this project at each stage of the process. The second article was co-authored by Paul Conway, Dan Bartels, and David Pizarro. All three authors contributed to the study concept. P. C. collected and analyzed the data and provided the first draft; D.M.B. and D.A.P. provided critical revisions. All authors approved the final version of the paper.

## Acknowledgments

This document reflects the culmination of my graduate school experience, during which I incurred massive debts to many parties.

First and foremost, I have boundless gratitude for my supervisor, Dr. Jim Olson, who has supported me in every conceivable way a supervisor can support their student. Jim really paved the path for my success through leading by example, and by providing feedback and encouragement, advice and guidance, financial support, and even caring for my cat Claire while I travelled to conferences in pursuit of my degree! As my academic father, Jim gave me the graduate school equivalent of secure attachment: he provided unwavering support as I toyed with the gamut of ideas in psychology, and any time I ran into difficulty I knew could return to him for advice and assistance that allowed me to overcome the obstacles in my way. It occurs to me that there is no way to repay a debt of this magnitude; instead it is incumbent upon me to *pay it forward*—rather than attempt to repay Jim for his kindness, I must treat my future students in the ways that he treated me. Hence, in all such matters I aim for the maxim WWJD (*What Would Jim Do?*). I hope he agrees that imitation is the sincerest form of flattery. As former student Chris Wilbur often said, “Is Jim amazing?”

I also owe a massive, unrepayable debt to Dr. Bertram Gawronski, who has long treated me as his student even though nothing ever required he do so. In classic Vygotskian fashion, Bertram “offered his mind to be read”—he reached down from his intellectual vantage point to scaffold my intellectual journey so that I became capable of mastering concepts beyond my individual unassisted capabilities. Now that I am armed with the information, techniques, and strategies learned from him, I feel well-placed to apply them to examine a range of fascinating questions pertinent to a wide range of issues. I cannot hope to repay such beneficence, but I can apply his lessons to push the horizon of psychological knowledge further forward. I hope I always reflect well upon him and his teachings.

So many others contributed to my success that it is difficult to acknowledge them all. I owe thanks to the members of my dissertation committee: Dr. Clive Seligman and Dr. William Fisher, whose insight and feedback improved this last, most important project of my graduate career, and who challenged me to broaden my perspective and consider elements that had not occurred to me. Thanks as well to Dr. Adam Cohen, Dr. Igor Grossman, and Dr. Charles Levin, who each offered insightful comments and criticism that helped me conceptualize this work in relation to a broader swath of thinking than I was previously able. Your comments have helped shape the trajectory of my research for years to come.

I also owe thanks to the other professors at Western, both within the social area and beyond it, from whom I have learned so much. Thanks to my collaborators who worked so tirelessly with me as we conceptualized, conducted, and crystalized the research herein. Thanks to the Social Sciences and Humanities Research Council of Canada for providing funding support for my research.

Thanks as well to my fellow graduate students who have been such great friends and colleagues through all the years. Graduate school has been the best part of my life, and I leave behind many cherished memories.

Thanks to my lab coordinators, research assistants, and honors students, who not only made this research possible, but who showed me that one of the greatest honors a scholar can have is to contribute to another scholar's achievement.

Finally, thanks to my family for supporting and enduring my decision to attend graduate school. It has been a long road, but I finally made it. Especial thanks to my Dad for first introducing me to psychology and showing me how it could lead one places. I'm glad to have followed in your footsteps.

# Table of Contents

Abstract .....	ii
Co-Authorship Statement.....	iv
Acknowledgments.....	v
Table of Contents .....	vii
List of Tables .....	x
List of Figures .....	xi
List of Appendices .....	xii
Chapter 1 .....	1
1 Introduction .....	1
1.1 The Original Trolley Dilemma .....	3
1.2 Theoretical Backdrop: Moral Cognition and Emotion .....	5
1.3 The Emergence of the Moral Dilemma Paradigm .....	11
1.4 The New Synthesis in Moral Psychology .....	15
1.4.1 Evolutionary Theory and the Dual Process Model .....	16
1.4.2 Cultural Variability in Moral Content.....	20
1.4.3 Universal Grammar Theory .....	22
1.5 Overview of the Field .....	26
1.6 A Problem: The Non-independence Error .....	29
1.7 Process Dissociation as a Potential Solution .....	32
1.8 The Current Project.....	41
References .....	44
Notes .....	58
Chapter 2.....	63



2	Deontological and Utilitarian Inclinations in Moral Decision-Making: A Process Dissociation Approach .....	63
2.1	Abstract .....	63
	Chapter 3 .....	65
3	Are Psychopaths Really More Utilitarian? Process Dissociation Clarifies the Relations Between Utilitarianism, Deontology, Anti-Social Personality Traits, and Prosociality.....	65
3.1	Abstract .....	65
3.2	Introduction.....	65
3.2.1	A Problem for the Field .....	66
3.2.2	Process Dissociation as a Solution.....	67
3.2.3	The Current Work .....	68
3.3	Study 1 .....	69
3.3.1	Method .....	69
3.3.2	Results and Discussion .....	70
3.4	Study 2 .....	72
3.4.1	Method .....	72
3.4.2	Results and Discussion .....	74
3.5	General Discussion .....	78
	References .....	81
	Notes .....	83
	Appendix 3A: Prosocial Dilemmas .....	87
	Chapter 4 .....	90
4	Conclusions General and Discussion .....	90
4.1	Results.....	90
4.1.1	Results: Chapter 2, Study 1 .....	90

4.1.2	Results: Chapter 2, Studies 2 & 3 .....	93
4.1.3	Results: Chapter 3, Studies 1 & 2 .....	94
4.2	Implications.....	98
4.2.1	Implications: Dual-Process Model.....	99
4.2.2	Implications: Moral Status of Utilitarianism .....	99
4.2.3	Implications: Interpreting Previous Results.....	101
4.3	Limitations .....	102
4.4	Future Directions .....	108
4.4.1	Gender Differences in Moral Judgments .....	108
4.4.2	The Role of Threat in Moral Judgments .....	109
4.4.3	Other Process Dissociation Projects .....	111
4.5	Summary .....	114
	References .....	116
	Appendix 4A: Copyright Permission.....	121
	Appendix 4B: Ethics Approval.....	122
	Curriculum Vitae <b>PAUL CONWAY</b> .....	127

List of Tables

Table 4: Regressing Utilitarian Decisions, PD Utilitarianism Scores, and PD Deontology Scores on Machiavellianism, Psychopathy, and No Meaning, Controlling for Social Desirability and Gender in Study 1 (N = 89)..... 84

Table 5: Regressing Utilitarian Decisions, PD Utilitarianism Scores, and PD Deontology Scores on Scenario Help, Volunteerism, Donations, Dilemma Help, help on Altruism, Honesty, and Charity Dilemmas, and Gender in Study 2 (N = 91). .... 86

## List of Figures

Figure 1: Processing tree illustrating the underlying components leading to judgments that harmful action is either acceptable or unacceptable in congruent and incongruent moral dilemmas. The paths from left to right depict the three cases that (1) utilitarianism ultimately drives the response, (2) deontology ultimately drives the response, and (3) neither utilitarianism nor deontology drives the response. .... 38

## List of Appendices

Appendix 3A: Prosocial Dilemmas .....	150
Appendix 4A: Copyright Permission .....	121
Appendix 4B: Ethics Approval .....	122

## Chapter 1

### 1 Introduction

Imagine that a vicious army commander captures your party as you travel through a remote region. If you kill one member of your party, this commander will allow the rest to go free; otherwise, he will kill everyone. Should you kill a member of your party to secure freedom for the rest? Or imagine you are a doctor performing surgery on a person with a rare blood type. If you kill this person and share their blood with other patients, you could save five lives. Should you kill this person in order to save the others? Or imagine you are captain of a capsized vessel and are now in an overcrowded lifeboat that is slowing sinking. If you push a grievously wounded person overboard, the rest of the people in the lifeboat will be safe; otherwise, everyone will drown. Should you push this person overboard in order to save the rest?

In essence, these questions entail causing some degree of harm in order to achieve a greater positive outcome. Philosophers have pondered such questions for hundreds of years, arguing over which actions are normatively correct (see Kagan, 1998). Some have argued that the morality of actions is defined by their nature rather than their consequences: some actions are intrinsically wrong and can never be right no matter what outcomes occur (e.g., Kant, 1785/1959). This perspective is called *deontology*, which means *duty* or *obligation* in Greek, and deontological ethical perspectives treat moral rules (such as do no harm) as absolutes that people are duty-bound to follow. According to deontology, causing harm is morally incorrect even in service of the best outcome, for doing so treats human beings as a means to an end, rather than an end in themselves. Kant argued that treatment as a mere means to someone else's end is fundamentally degrading; therefore, causing harm is morally unacceptable regardless of outcomes.

Others have argued that harm dealt in the process of helping many others is morally permissible—nay, morally mandated—because the morality of an action should be judged according to the consequences that action produces (e.g., Mill, 1861/1998).

These perspectives are generally referred to as *consequentialist* ethical positions, as they are focused on the consequences of actions, but the best-known variety is *utilitarianism*, which mandates acting in ways that maximize overall *utility* (i.e., well-being) for all parties concerned (Kagan, 1998). According to utilitarian perspectives, actions that produce net positive outcomes are morally acceptable and mandated, even if action results in harm (provided that more utility is gained than lost by harmful action).

Both deontology and utilitarianism are described as moral principles. Defining morality is incredibly difficult given the vast scope of the construct, so I will present several recent definitions in an effort to be thorough. In a review chapter, Krebs (2008) defined morality as a “mental phenomenon that consists in thoughts and feelings about rights and duties, good and bad character traits (virtues and vices), and right and wrong motives and behaviors” (p. 150). In another review chapter, Haidt and Kesebir (2010) expanded this definition to consider elements of morality outside of psychological perception: “Moral systems are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make cooperative social life possible” (p. 800). In yet another review, Skitka, Bauman, and Mullen (2008) noted that morality entails perceptions of *universality* (everyone must agree that a judgment is morally correct; disagreement is taken as evidence of immoral character), *objectivity* (moral judgments are perceived as objectively true similar to factual information about the world, such as  $2 + 2 = 4$ ), and *authority independence* (moral rules do not depend on authority; if an authority acts against moral rules this abrogates their authority rather than the rules). Moreover, morality entails a motivational component: upon making a moral judgment, people feel compelled to act on their judgment, lest they view themselves as a hypocrite (Skitka, Bauman, and Mullen (2008). Finally, morality is linked to identity and self-perceptions, although there is variance in the degree to which morality is central to the self-concept (Aquino & Reed, 2002; Hardy & Carlo, 2005; Bergman, 2002). Thus, both deontology and utilitarianism theoretically qualify as genuinely moral

insofar as they entail a concern with others' wellbeing, which the various definitions agree is a key element of morality.

## 1.1 The Original Trolley Dilemma

In most real-world situations, deontological and utilitarian ethics lead to the same judgment. Actions such as *theft*, *lying* (for selfish gain), and *murder* violate both the deontological principle to avoid harm and the utilitarian ideal of maximizing overall outcomes (except in rare cases, such as stealing something to save a life). Therefore, following either ethical perspective ought to lead to morally acceptable conduct in most situations—which means that moral conduct in most situations remains uninformative regarding which moral principle actors ascribe to. Yet, it is possible to imagine rare circumstances where deontological and utilitarian perspectives conflict. Philosopher Philippa Foot (1967) introduced one such scenario as a thought experiment:

“He is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed... The question is why we should say, without hesitation, that the driver should steer for the less occupied track....” (p. 8-9).

Foot was contrasting this trolley example with that of a corrupt magistrate who frames an innocent man in order to stop a riot, and she argued that the former is acceptable whereas the latter is not due to the *doctrine of double effect*, whereby a harmful action ought to be viewed differently depending on whether harm was dealt *in order* to achieve a positive outcome versus harm dealt as a *side effect* of another action aimed at achieving a positive outcome (Aquinas, 1265-1272/1947). She argued that harm in the trolley case is acceptable, as the harm will only occur as a side-effect of acting to maximize outcomes (i.e., changing the path of the train). Conversely, in the case of the magistrate, harm to the framed innocent is the *mechanism* through which outcomes are to be maximized. Therefore, Foot argued that this kind of harm is unacceptable.



Foot's distinction is interesting, but subsequent work has not examined this distinction in detail (cf., Manfrinati, Lotto, Sarlo, Palomba & Rumiati, 2013). Rather, Foot's trolley problem (or tram problem, as she may have viewed it) became widely known because it illustrates a rare tension between deontological and utilitarian ethical positions. Essentially, the trolley problem asks whether is it morally acceptable to kill one person (or knowingly allow one person to be killed) in order to save five people. According to deontological ethical positions, the morally correct answer is *no*—dealing harm is unacceptable even though it will save lives, because this harm treats the one person as a means to the ends of the other people; treating people as means to an end is fundamentally degrading. People have a moral duty to treat one another as ends in themselves (Kant, 1785/1959). According to utilitarian ethical positions, the morally correct answer is *yes*—in this case, causing harm is a lesser evil that must occur in order to achieve the greater good, and is therefore both morally permissible and morally mandated (Mill, 1861/1998). As the good achieved by the action (five lives saved) is greater than the evil achieved (one life ended) this action is considered morally correct by utilitarian standards. Therefore, utilitarianism and deontology mandate different actions in moral situations where causing harm is necessary in order to achieve the greater good.

For over a decade now, psychologists, neuroscientists, and other researchers have moved beyond thought experiments to examine the empirical distribution of lay judgments on such moral problems, particularly following Greene and colleagues' landmark publication in *Science* (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Whereas philosophers had been largely content to speculate regarding how people *should* answer such dilemmas, researchers began to administer trolley-type moral dilemmas to various populations and examined the empirical distribution of how people *actually answer* moral dilemmas, in an effort to improve understanding of human moral psychology (e.g., Bartels, 2008; Carney & Mason, 2010; Côté, Piff, & Willer, 2013; Conway & Gawronski, 2013; Ciaramelli, Muccioli, Ladavas, & di Pellegrino 2007; Fumagalli et al., 2010; Greene et al., 2001; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Hofmann & Baumert, 2010; Koenigs et al., 2007; Manfrinati et al., 2013;

Mendez, Anderson, & Shapria, 2005; Moore, Clark, & Kane, 2008; Nichols, 2002; Nichols & Mallon, 2006; Pellizzoni, Siegal, & Surian, 2010; Petrinovich & O'Neill, 1996; Petrinovich, O'Neill, & Jorgensen, 1993; Valdesolo & Desteno, 2006).

Although researchers have debated the normative implications of empirical findings in this domain (e.g., Baron, 2000; Baron & Spranca, 1997; Bartels & Pizarro, 2011; Bazerman & Greene, 2010; Benis, Medin, & Bartels, 2010; Greene, 2003; Sunstein, 2005), most of the debate has shifted away from normative discussion to discussion of mechanism: what are the psychological processes that motivate judgments that harm is acceptable or not acceptable on such dilemmas? We will return to this discussion momentarily, but first let's consider the broader picture regarding why scientists were examining responses to moral dilemmas in the first place.

## 1.2 Theoretical Backdrop: Moral Cognition and Emotion

As previously mentioned, cases where deontology and utilitarianism conflict are rare indeed: how often does one have an opportunity to save several lives, let alone sacrifice another life in order to do so? Indeed, from an outside perspective, it might seem odd that scientists study the question at all—for the aim of this research is not to give people practical advice on whether or not to sacrifice another person under particular circumstances. Rather, this question is of interest because it provides evidence regarding the fundamental nature of human moral psychology as a whole—a field that was undergoing a Kuhnian (1989) revolution when Greene and colleagues' (2001) paper was published.

Since its inception, the field of moral psychology had been dominated by the theorizing of two influential 20<sup>th</sup> Century scholars: Jean Piaget and Lawrence Kohlberg. Piaget (1932/1965) described the ontogenetic development of children's thinking in terms of logic structures: he argued that children advance in intelligence by integrating successively more abstract information into coherent concepts across hierarchical stages of development. Piaget argued that children begin life without moral concepts, and then develop first *heteronomous* moral reasoning (moral rules are those enforced by authority) and then *autonomous* moral reasoning (moral rules are negotiated fairly between equals). Kohlberg (e.g., 1969) expanded Piaget's model to

encompass six reasoning stages of moral reasoning (for a review, see Kohlberg & Hersh, 1977).

*Moral reasoning* is “conscious mental activity that consists of transforming given information about people in order to reach a moral judgment” (Haidt, 2001, p. 818). In other words, moral reasoning is at least partially intentional, effortful, controllable, and occurs within the conscious awareness of the reasoner (Bargh, 1994), and involves sequential steps (Galotti, 1989) of constructing and testing hypotheses (Turiel, 1983) by searching for and weighing evidence (Kuhn, 1989) pertaining to harm, justice, fairness, motives, and other aspects of the situation. Accordingly, Kohlberg measured reasoning by coding the breadth and sophistication of participants’ verbal responses to various ethical quandaries (e.g., *should a man steal a drug to save his wife’s/friend’s/a stranger’s life?*). He argued that increasing sophistication in the quality of one’s moral reasoning is both the empirical cause of moral judgments and behavior, as well as the normative, teleological goal of ontogenetic development (Kohlberg, 1971).

According to Kohlberg, morality is inherently rational: *to know the good is to do the good*. No individual can claim moral knowledge without the simultaneous realization that this knowledge requires them to act morally—because moral principles are universal and prescriptive categorical imperatives that demand action (Kohlberg & Candee, 1984). Indeed, moral imperatives are more than just strong attitudes (see Krosnick & Petty, 1995), they are experienced as factually correct beliefs about the world (Shweder, 2002) that ought to be universally adopted (Turiel, 1983), and they provide both motivation and justification for action (Skitka, Bauman, & Sargis, 2005). Therefore, Kohlberg argued that the primary reason people fail to engage in moral behavior is because they have failed to engage in (sufficiently developed) moral reasoning. Kohlberg dismissed “irrational emotive theories” as exogenous to human morality (Kohlberg, 1971, p. 188). Emotions are neither moral nor immoral, he argued, since they can be channeled toward either moral or immoral ends. The channeling processes themselves are cognitive reasoning structures based upon rational, verbal thought. This view aligns with a rich intellectual tradition including

Immanuel Kant (1785/1959), and Plato (1949), who were skeptical that ‘the passions’ should be considered part of human morality.

Kohlberg’s perspective proved so popular that it went on to define an entire generation of psychological research, called *cognitive developmentalism* (see Blasi, 1980, and Lapsley, 1996, for reviews). Although cognitive developmental scholars critiqued Kohlberg in numerous ways, nonetheless they unanimously agreed with Kohlberg’s central premise that reasoning drives moral judgment. For example, Gilligan (1982) argued for expanding the moral domain to include a more ‘feminine’ ethic of caring alongside the ‘masculine’ ethic of justice. However, she utilized the Kohlbergian moral judgment interviews to argue her case. Likewise, Turiel (1983) borrowed Kohlberg’s methodology even as he demonstrated that participants distinguish the moral domain from social convention and personal preferences. Rest (1986) argued that Kohlberg’s cognitively deterministic model oversimplified the complex process of moral action, and that other factors (including empathy) ought to be considered. He argued to incorporate the work of Hoffman (1982), who emphasized the role that moral emotions play in moral judgments, with that of Kohlberg. Nonetheless, Rest’s (1979) Defining Issues Test focused primarily on the same rational decision-making processes that Kohlberg’s stages did (Bergman, 2002). Even Blasi, who argued that moral reasoning fails to motivate moral behavior unless the self is involved, still believed that only reasoning allows for integrating morality into identity, since it requires “personal effort to sift various opinions” (Blasi, 1995). Thus, Blasi accepted the role of rationality in making moral judgments, regardless of the import of the self in moral action. In sum, despite decades of progress and numerous challenges, the field of moral psychology was dominated by rationalist models of judgment right up into the 1980s and 1990s. Yet, this state of affairs would not last.

At the turn of the new millennium, Haidt (2001) published a landmark paper calling into question the assertion that moral reasoning drives (most) moral judgments. Instead, Haidt argued that moral judgments are the product of *moral intuitions*: “the sudden appearance in consciousness of a moral judgment, including an affective

valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion” (Haidt, 2001, p. 817). Haidt argued that intuitions are distinct from reasoning in that intuitive judgments occur quickly and effortlessly, in a one-shot process that occurs largely, if not entirely, outside of conscious awareness (Bargh, 1994). Whereas the *judgment* itself may enter into conscious awareness, the *process* of arriving at that judgment does not. Moreover, Haidt argued, moral reasoning (often) operates post hoc—reasoning simply justifies the judgment caused by intuitions, like a lawyer defending her client’s position regardless of why her client is in the legal system.

Haidt did allow that moral reasoning may sometimes directly influence judgments, but he argued this would occur infrequently (others are more optimistic about the influence of reason on intuition, see Pizarro & Bloom, 2003; cf. Haidt, 2003b).

Rather, he argued that the primary purpose of moral reasoning is *social*—to communicate one’s moral judgments to others in ways that hopefully evoke the same intuitions in them, thereby achieving social consensus regarding moral judgments (although this process may also result in deadlocked arguments between opposing moral camps, such as in the abortion debate, Graham, Haidt, & Nosek, 2009). This argument places Haidt in a venerable intellectual tradition including Aristotle (1962) and David Hume, who said that “reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (1739/1969, p. 462).

The social intuitionist model posed a serious challenge to rationalist models of moral psychology, in part because it resonated with a host of findings casting doubt on the predominance of rational thinking overall. Evidence suggests that people are often unaware of the causes of their behaviour (Nisbett & Wilson, 1977) as much processing takes place below the threshold of conscious awareness (Greenwald & Banaji, 1995). As cognitive misers, humans often engage in heuristic processing unless there is need for more systematic processing (Chaiken, 1987). Once people have arrived at an initial conclusion, they tend to reinforce it by seeking confirmatory

evidence (Kunda, 1990)—particularly when motivated to do so, as when defending a moral position (Kuhn, 1989).

People make affect-laden judgments of facial (Murphy & Zajonc, 1993) and verbal (Bargh, Chaiken, Raymond, & Hymes, 1996) stimuli within 250 milliseconds of subliminal presentation. People also categorize others into morally valenced stereotypes within milliseconds (see Devine, 1989). People rely on affective information when making justice judgments (Van den Bos, 2003). Luo and colleagues (2006) found negative evaluations of immoral stimuli on an implicit association test were associated with activation in affective brain areas. In addition, the motivational force of moral judgments is mediated by emotion (Mullen & Skitka, 2006). The particular moral emotion experienced may depend on which domain of morality has been violated (e.g., harm vs. purity violations may elicit more anger vs. disgust, Rozin, Lowery, Imada, & Haidt, 1999; Cushman, Young, & Hauser, 2006; Shweder, Much, Mahapatra, & Park, 1997; Ugazio, Lamm, & Singer, 2012). That said, it may not always be necessary to hone in on the exact moral emotion in play, as the moral emotions of anger and disgust are positively correlated, and interact to predict moral judgments (Salerno & Peter-Hagene, 2013, cf., Russel & Giner-Sorolla, 2011).

Setting aside questions regarding specific emotional states, evidence suggests that when people encounter a morally charged stimulus, they experience affective reactions that predict their moral judgments better than do their rational arguments (Haidt, Koller, & Dias, 1993)—unless they actively suppress their emotional response in order to avoid the behavioral implications of moral emotions (Cameron & Payne, 2011). Accordingly, the experience of affective states increases the intensity of moral judgments (Horberg, Oveis, & Keltner, 2011). Merely simulating actions that would cause harm elicits visceral, physiological reactions, especially for people higher in threat reactivity (Cushman, Gray, Gaffey, & Mendes, 2012). Participants view moral judgments regarding intentional harm as universal in scope and independent of authority wishes (Sousa, Holbrook, & Piazza, 2009). Moreover, moral emotions (not

reasoning or identification) carry the motivational force in moral judgments (Skitka, Bauman, & Sargis, 2005).

Moreover, individual differences in the propensity to experience moral affect predict guilt and reactions to injustice (Hofmann & Baumeister, 2010), and individual differences in sensitivity to disgust predict a wide variety of moral judgments and behavior, including disapproval of gay people (Inbar, Pizarro, Knobe, & Bloom, 2009), moral attitudes regarding animal welfare (Herzog & Golden, 2009), condemnation of unusual sexual behaviour (Haidt & Hersh, 2001), and even perception of morally relevant perceptual stimuli (Sherman & Clore, 2009). Experimentally inducing disgust increases the severity of moral judgments (Schnall, Haidt, Clore, & Jordan, 2008). This effect does not require awareness of the link between disgust and moral judgments—Wheatley and Haidt (2005) found that participants were more likely to judge actions as morally incorrect when disgust was hypnotically induced versus a control group. In fact, disgust induction was so effective in this study that one third of participants rated a *neutral* story as somewhat morally wrong even though they could not verbalize a reason for doing so.

But perhaps the most striking evidence for the social intuitionist model comes from a study conducted by Haidt, Koller, and Dias (1993). They constructed stories in which the protagonists behaved offensively, but failed to cause any real harm. Participants' reactions to these stories of incest, having sex with a chicken, or eating one's pet dog tended to be strongly negative; most participants immediately and adamantly asserted that such actions were morally wrong—yet they struggled to find rational reasons why they believed so. Haidt and colleagues coined the term *moral dumbfounding* to describe this perplexing state of affairs. If moral judgments really are the product of moral reasoning, and reasoning is at least partially conscious, then participants ought to be able to articulate the reasons behind their judgment. Findings such as those of Haidt et al. (1993) led some theorists to even claim that “emotions are not merely correlated with moral judgments but they are also, in some sense, both necessary and sufficient” (Prinz, 2006, p. 29). Yet, the social intuitionist model has also faced

criticism (e.g., Narvaez, 2008; Monin, Beer, & Pizarro, 2007; Saltzstein, & Kasachkoff, 2004).

In sum, Haidt's (2001) social intuitionist model challenged older rationalist models of moral psychology by proposing that moral judgments primarily result from fast, visceral, affective moral intuitions, and that moral reasoning primarily operates as post-hoc justification. This position stood in stark contrast to the Kohlbergian (e.g., 1971) view that morality stems only from rational processes. Hence, the field was characterized by two fairly extreme and largely incompatible perspectives that disagreed over the primacy of affect versus cognition. Each perspective labeled one process as primary and the other secondary. The field was mired in dissention, until a new paradigm emerged, allowing researchers to pose new questions regarding human moral psychology.

### 1.3 The Emergence of the Moral Dilemma Paradigm

In the same year that Haidt published his influential social intuitionist model, another group of scholars led by Joshua Greene published what became a landmark paper in *Science* (Greene et al., 2001). They argued that *neither* cognition nor visceral, affect-laden constructs were primary in moral psychology; rather, each process motivates *different kinds* of moral judgments. Specifically, affect-laden processes motivate moral judgments to avoid causing harm, whereas deliberative reasoning motivates moral judgments to maximize welfare.<sup>1</sup> In contrast to other theoretical positions (i.e., Haidt, 2001; Kohlberg, 1971), this distinction provided a framework wherein both affect-laden processes and rational deliberation may causally impact moral decision-making. Indeed, Greene and colleagues claimed their model “allows us to steer a middle course between traditional rationalism and more recent emotivism that have dominated moral psychology” (2001, p. 2107).

According to Greene and colleagues' model, when faced with a moral quandary where one person must be hurt in order to aid a number of others, people immediately and involuntarily experience a negative, affect-laden reaction to the prospect of causing harm. If this affective reaction is sufficiently powerful, or in the absence of



sufficient time, motivation and resources to engage in deliberation, it will dominate the decision-making process, resulting in a judgment that harm is morally unacceptable (consistent with deontological ethical positions). Under more generous processing conditions that allow for complex cognition, people may also engage in cognitive deliberation regarding the costs and benefits of harming another person. Given sufficient time, motivation, and resources, cognition will dominate decision-making, resulting in a judgment that harm is morally acceptable (consistent with utilitarian ethical positions).<sup>2</sup> Thus, Greene and colleagues proposed that the psychological processes underlying deontological and utilitarian moral decisions are distinct and potentially independent, making conflict possible. Greene argued this dual-process model parallels dual-process models in other domains, such as stereotyping (Devine, 1989).

Like all good models, the dual-process model of moral judgments suggests novel, testable hypotheses regarding the conditions where one process will predominate over the other. Contexts where moral stimuli elicit strong affective reactions or where deliberation is difficult—such as physical proximity to the victim or time pressure—should promote judgments based on visceral, affective, or heuristic processes, thereby increasing the likelihood of judging harm unacceptable (consistent with deontological ethical positions). Conversely, in situations that fail to evoke strong affect, or that are conducive to deliberation—such as physical distance from the victim or time to deliberate—people should be more likely to cognitively evaluate outcomes, thereby increasing the likelihood of making judgments that harm is acceptable (consistent with utilitarian ethical positions). Moreover, individual differences in cognitive abilities, such as working memory capacity, should be associated with more utilitarian judgments, whereas individual differences in affective tendencies (e.g., empathic concern) should be associated with more deontological judgments.

Greene and colleagues (2001) provided initial evidence for the theory using both self-reports and functional magnetic resonance imaging (fMRI) data obtained by posing moral dilemmas to participants in a scanner. Like Foot (1967), Greene contrasted different dilemmas to make his point. Unlike Foot, however, Greene did not focus on

the doctrine of double effect (the distinction between achieving positive outcomes via harmful actions directly, versus achieving positive outcomes via methods that cause harm as an incidental byproduct of action).<sup>3</sup> Instead, Greene followed philosopher Thompson (1986) by contrasting responses on dilemmas where harmful actions are completed through a technical intermediary, such as pushing a button to make a train change direction, which he called *impersonal dilemmas*, with similar dilemmas where harm is applied through direct contact with the victim, which he called *personal dilemmas* (although subsequent researchers have typically renamed this variant *high-conflict dilemmas*, e.g., Koenigs et al., 2007, or *incongruent dilemmas*, e.g., Conway & Gawronski, 2013). For comparison, Greene and colleagues also examined nonmoral dilemmas, such as choosing whether to travel by bus or train, but we will focus discussion on the personal-impersonal distinction.

Greene's version of the impersonal trolley dilemma reads as follows:

"A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Ought you to turn the trolley in order to save five people at the expense of one?" (Greene et al., 2001, p. 2105).

Conversely, Greene's version of the personal footbridge dilemma reads:

"As before, a trolley threatens to kill five people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the five people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Ought you to save the five others by pushing this stranger to his death?" (Greene et al., 2001, p. 2105).

Greene and colleagues asked participants whether causing harm was *acceptable* or *not acceptable* in each case. On personal dilemmas similar to the footbridge dilemma, most people (~80%) selected the deontological option (i.e., harm is not acceptable). Conversely, on impersonal dilemmas similar to the trolley dilemma, most people (~80%) selected the utilitarian option (i.e., harm is acceptable). This basic pattern has now been replicated many times (e.g., Greene et al., 2004, Carney & Mason, 2010), and even emerges for children as young as three (Pellizzoni et al., 2010).

Greene and colleagues argued that both personal and impersonal dilemmas pit deontological considerations against utilitarian ones, but these dilemma variants differ to the extent that they elicit affective reactions to harm. On personal dilemmas, participants must contemplate harm they cause directly, thereby enhancing the salience of that harm and the visceral, emotion-laden reaction against causing it. On impersonal dilemmas, harm is mediated by technical devices and therefore less salient—purportedly reducing the strength of affective reactions, and allowing cognition to shine through. As Greene and colleagues (2001) put it, “The thought of pushing someone to his death is, we propose, more emotionally salient than the thought of hitting a switch that will cause a trolley to produce similar consequences, and it is this emotional response that accounts for people’s tendency to treat these cases differently” (p. 2106).

Subsequent work clarified that the feature of personal dilemmas that promotes affective responses is *personal force*: intentionally using one’s physical strength to directly cause harm (as in pushing someone off a bridge, Greene et al., 2009). Hence, pushing someone off a bridge with a long pole typically engenders responses that harm is not acceptable (as the force originates in one’s muscles), whereas pressing a button to have a person fall through a trap door directly in front of you typically engenders responses that harm is acceptable (as the force of one’s muscles is transformed via mechanical action). Therefore, it appears that using personal force to intentionally harm another person elicits stronger reactions than using mechanical intermediaries to cause harm. Yet, other research has demonstrated that a variety of techniques are effective at increasing judgments that harm is unacceptable, such as

visualizing harm more vividly, responding to dilemmas written in an emotionally evocative manner, or responding to dilemmas accompanied by visual depictions of the victim (Amit & Greene, 2012; Bartels, 2008; Conway & Gawronski, 2013). These findings suggest that personal force is but one of many manners in which the emotional salience of causing harm may be highlighted.

In addition to self-report data on whether participants viewed harm as acceptable or unacceptable, Greene and colleagues also examined patterns of activation in various brain regions via fMRI.<sup>4</sup> They found that emotion centers in the brain, such as the medial prefrontal cortex, posterior cingulate, and bilateral superior temporal sulcus, evinced greater activation when participants considered personal compared to impersonal moral dilemmas (Greene et al., 2001). A replication by Greene and colleagues (2004) found activation in most of the same brain regions, but added the amygdala to the list of brain regions activated when participants considered personal dilemmas. Moreover, Greene and colleagues (2004) examined activation patterns when participants made deontological judgments that harm is not acceptable—largely the same regions demonstrated activation.

Greene and colleagues also found that when participants considered impersonal rather than personal dilemmas, there was greater activation in brain regions linked to abstract reasoning, working memory, and conflict resolution, such as the right dorsal lateral prefrontal cortex and inferior parietal lobes (Greene et al., 2001), as well as the anterior cingulate cortex and the posterior cingulate cortex (Greene et al., 2004). These same regions show increased activation when participants made utilitarian judgments (i.e., judged harm as acceptable on a dilemma). They argued that these findings provided initial evidence in support of the dual-process model where affective reactions to harm drive deontological judgments and cognitive evaluations of outcomes drive utilitarian judgments.

## 1.4 The New Synthesis in Moral Psychology

Haidt (2007) hailed Greene and colleagues' insight regarding the role of affect and cognition in morality as part of *The New Synthesis in Moral Psychology*—a synthesis

that incorporates insights from neuroscience (Moll, de Oliveira-Souza, & Eslinger, 2003), evolutionary theory (Darwin, 1874/1998), primatology (de Waal, 1996), and social psychology (e.g., Zajonc, 1980) to clarify current understanding of the nature of human moral psychology. Haidt argued the new synthesis suggests:

“People are self-interested, but they also care about how they (and others) treat people, and how they (and others) participate in groups. These moral motives are implemented in large part by a variety of affect-laden intuitions that arise quickly and automatically and then influence controlled processes such as moral reasoning. Moral reasoning can correct and override moral intuition, though it is more commonly performed in the service of social goals as people navigate their gossipy worlds” (Haidt, 2007, p. 1001).

Although reasoning is part of the equation here, the new synthesis does not include the intellectual traditions of Kohlbergian moral stage typology (see Haidt & Kesebir, 2010). Instead, the modern view of moral cognition centers on the concept of weighing the costs and benefits of action—i.e., utilitarian tradeoffs—rather than a stage-structure of moral reasoning. Haidt argued that moral dilemma research resolves the tension between intuitive and rational approaches to moral psychology by delineating how and when each process impacts moral judgments.

#### 1.4.1 Evolutionary Theory and the Dual Process Model

Although not designed to test evolutionary claims, the dual-process model of moral psychology is consistent with modern evolutionary theory. Greene and Haidt (2002) argued that affective reactions to harm may stem from evolutionarily ancient psychological mechanisms designed to inhibit aggression and promote prosociality among social species, so that they may reap the benefits of cooperation. They also argued that cognitive evaluations of outcomes, as a deliberative process, are tied to language and domain-general abstract reasoning, and are therefore likely more evolutionarily recent. Yet, the uniquely human capacity to reason linguistically may have facilitated negotiation between social entities on a more abstract scale, allowing

for wide-scale social cooperation orders of magnitude more complex than other social species.

Similar arguments have been made by theorists delineating an evolutionary account of human morality. As morality generally entails acting in ways that promote the welfare and wellbeing of others, typically at some personal cost, it might seem antithetical to evolutionary principles. Paradoxically, however, moral behaviors may have increased the inclusive fitness of individuals who performed them—at least under certain conditions. According to Trivers' (1971) theory of reciprocal altruism, altruistic behaviors will be selected for when (1) the cost to the actor is smaller than the benefit to the recipient, and (2) there is a high probability that the favor will be returned. Because both interaction partners suffer small costs relative to benefits they receive, both profit from a net fitness increase. Total costs subtracted from total benefits net a sum larger than zero, leading theorists to term such interactions *non-zero-sum* (Wright, 1994).

Although Trivers envisioned reciprocal altruism as a one-on-one interaction, non-zero-sum fitness benefits increase as reciprocity relationships expand to encompass a cooperative group (Wright, 1994). Members of such groups incur small fitness costs (i.e. act morally) in order to provide larger fitness benefits to other group members. They, in turn, benefit from the sacrifices of other group members. The system operates as one of *indirect reciprocity* (Alexander, 1989), which Darwin (1871) termed the 'foundation-stone' of morality. The indirect reciprocity model meshes well with empirical data on the environment of human evolution. For the last several million years, humans lived together in small, highly interdependent, stable hunter-gatherer groups (Lee & Devore, 1968). These groups meet all the theoretical conditions required to make systems of indirect reciprocity increase fitness (Burnstein, Crandall, & Kitiyama, 1994). Moreover, under these conditions, group selection processes (Sober & Wilson, 2001) as well as sexual selection pressures (Hardy & Van Vugt, 2006; Miller, 2007) may all contribute to the selection of psychological mechanisms that motivate prosocial behavior. Krebs (2008) pointed out that once selection pressure existed to act in prosocial ways, it might lead to cases of

genuine, genetic altruism—heroic sacrifice of one’s individual genetic fitness for greater good in a way that does not indirectly foster genetic fitness—so long as *on average* those mechanisms served to increase inclusive fitness. Hence, “in evolutionary theory, a moral person is simply one who pursues their ultimate genetic self-interest through psychological adaptations that embody a genuine, proximate concern for others” (Miller, 2007, p. 103).

Accordingly, people have a tendency to cooperate in single-round prisoners dilemmas even when it is logically advantageous to cheat (Macy & Skvoretz, 1998) and are willing to punish a cheater who hurt someone else, even at personal cost (Gintis, Bowles, Boyd, & Fehr, 2003). These costly behaviors only make sense if heuristically cooperating contributes to the fabric of the social group, thereby indirectly increasing fitness outcomes for oneself through maintaining cooperative norms. In reciprocal exchange, strategies such as tit-for-tat that cooperate with other altruists but defect with other cheaters, are evolutionarily stable and capable of invading populations comprised of either undifferentiating co-operators or cheaters (Axelrod & Hamilton, 1981). However, it also makes sense to heuristically reject those who fail to cooperate; this simple strategy (with a few modifications) of cooperating with others *unless they fail to uphold their end of the bargain* (i.e., punishing free-riders) is capable of becoming evolutionarily stable when pitted against other strategies in computer simulations (Axelrod & Hamilton, 1964). Gintis et al, (2004) demonstrated that so long as humans in cooperative systems were able to track others’ reputations and sanction one another cooperative systems flourished; when people lost the ability to sanction cooperation plummeted.

It is important to note that although reciprocity relationships greatly enhance inclusive fitness, they also hinge on a delicate balance of interests. Insofar as individuals profit from reciprocating with one another their interests coincide; however, because of genetic unrelatedness, this overlap is not complete (Alexander, 1987). Therefore, selection will favour individuals who “cheat” the system by contributing less than their peers (Trivers, 1971). Cheating spells disaster for a system of indirect reciprocity. Cheaters receive the same level of fitness benefits as do

reciprocators, but at a reduced cost, and therefore have higher inclusive fitness. This fitness advantage will cause genes that facilitate cheating to spread through a cooperative system until indirect reciprocity ultimately breaks down (Dawkins, 1976).

In order to prevent this eventuality, systems of indirect reciprocity require rules regarding relative contributions and benefits: a moral code. Since altruism increases fitness only under certain conditions, there would have been selection pressure for behavioural flexibility (Gangestad et al., 2006). Evolved psychological mechanisms that prescribe altruism should be sensitive to the conditions when altruism increased fitness and when it did not, leading to situationally appropriate behaviour (Krebs, 2005). This assumption is in line with current theorizing on the nature of evolved mechanisms: they operate as “if-then” decision-rules or strategies (Buss, 2004), that dispose individuals to respond differently in different social contexts (Barkow et al., 1992).

People and animals appear to accurately track the costs suffered and benefits enjoyed by each group member, and tend to adjust behavior accordingly, maintaining reciprocity (Trivers, 1971; Flack & de Waal, 2001). People are extremely adept at detecting cheaters (Cosmides & Tooby, 2005). Those caught cheating arouse indignation; indignant individuals employ a variety of sanctions called *moralistic aggression* (Trivers, 1971). Accordingly, attitudes research indicates that emotion carries the motivational force in moral judgments—and not just any behavior, but strategic behavior that enforces moral norms. For example, gratitude motivates repayment of social debts (McCullough, Kilpatrick, Emmons, & Larson, 2001), guilt motivates repair of damaged relationships (Baumeister, Stillwell, & Heatherton, 1994), and anger motivates aggression at moral violators to sanction them for their violation (Skitka, Bauman, & Sargis, 2005). Thus moral emotions may serve to motivate strategic behaviors that enable the system of reciprocity to endure, providing a net increase in inclusive fitness for all group members.



Essentially, moral judgments and behaviors seem geared towards upholding the social contract (Flack and de Waal, 2001). Individuals agree to contribute to the group (act morally) in return for the assurance that everyone else will. Violation of this contract (immoral behaviour) carries swift and stiff penalties (moralistic aggression).

Morality, therefore, enables genetically selfish organisms to come together and harvest the fruits of non-zero-sum exchanges (Alexander, 1987), making group life possible (Flack and de Waal, 2001). When protected by morality, indirect reciprocity becomes an effective, attractive strategy. Genes that foster psychological mechanisms that foster it may out-compete those that give rise to relentless Hobbesian competition. Thus, it seems reasonable to posit that people have ancient, affect-laden responses to particular stimuli as those particular stimuli have repercussions for fitness outcomes, and therefore have played an important role in human evolutionary history. This may be why the tendency to develop and enforce moral codes is universal across cultures (Midgley, 1991).

#### 1.4.2 Cultural Variability in Moral Content

It is worth noting that although the tendency to develop and enforce moral codes is universal, there also remains considerable variability in moral judgment. This is not a paradox—variation is part of evolution, and contexts can also shape moral norms by influencing which evolved strategies are most lucrative (for a discussion, see Haidt & Joseph, 2004). To account for this, Rozin and colleagues (Rozin, Lowery, Imada, & Haidt, 1999) proposed the ‘CAD hypothesis’ that in some cultures (e.g. India) morality was about more than just harm to victims (violations of *autonomy*)—violations could also harm the moral domains of *community* and *divinity* (Haidt, Koller, & Dais, 1993). Whereas the morality of harm focuses on the right of each individual to be free of infringement/harm by his neighbour, the morality of community focuses on how well members of society fulfill the social role they are allotted. Dutifully playing one’s role contributes to a smooth functioning society, which is of value to everyone. The morality of divinity focuses on the purity of body, mind, and soul, and avoiding sinful acts that debase the temple of the body. Each of these three domains of morality has been linked to specific other-directed moral

emotions: violations of autonomy generate anger, violations of community generate contempt, and violations of purity generate disgust. Haidt and Graham (2007) broadened this distinction into five moral foundations (Harm, fairness, loyalty, respect for authority, and purity), and Haidt (2012) has since added a sixth foundation, freedom.

According to Rozin's classification system, moral dilemmas deal only with autonomy violations; they do not examine issues of purity or community. According to Haidt's classification system, moral dilemmas involve a clash between harm and fairness foundations—people wish to avoid causing harm (deontology), but also wish to be fair and rational, which might entail utilitarian sacrifice. They do not touch on moral issues pertaining to purity violations (such as the disgusting but harmless stories employed by Haidt et al, 1993) or violations of loyalty or authority. Thus, although moral dilemma research may be informative regarding some kinds of moral violations, it does not explain all of moral psychology; it is possible there are other systems involved in moral violations of purity or loyalty.

Haidt (2008) argued that psychologists have traditionally examined a limited scope of the moral domain, in part because most scientists are White, Educated, liberal, and etc. Moral dilemma research may be vulnerable to that criticism in that it examines only harm and fairness violations; it is silent regarding the psychology of violations in other moral domains. Although there is cultural variability in terms of the importance of moral foundations Haidt et al, 1993)—as well as liberal-conservative differences—it is worth noting that nearly all participants in large international samples moralize the domains of harm and fairness (Graham, Haidt, & Nosek, 2009). Thus, moral dilemma research examines issues that tend to be perceived as morally relevant to a wide variety of people no matter what their cultural background.

In sum, there are reasons to believe that visceral, affect-laden reactions to particular, morally-relevant stimuli may ultimately stem from pressures in the deep evolutionary past, which shaped psychological mechanisms that motivated strategic behaviors that, heuristically, on average, may have improved fitness outcomes in a given context. In

addition, the human ability to engage in abstract, symbolic, deliberative, language-based reasoning may have also been a product of more recent evolutionary processes, enabling humans to weigh many considerations against one another and make moral judgments about abstract concepts beyond the visceral, face-to-face morality of affect-laden moral judgments. If so, then both affective and cognitive moral processing may have an evolutionary origin; at least, the dual-process theory is compatible with such a view. However, the dual-process model is not the only theory of moral judgments consistent with evolutionary theory; another theory vies to explain moral dilemma judgments.

### 1.4.3 Universal Grammar Theory

Universal Moral Grammar theory (Mikhail, 2007) is modeled on Chomskyan linguistic theory (e.g., Chomsky, 1964), and aims to integrate computational, ontogenetic, behavioral, physiological and phylogenetic elements of morality. It entails two crucial arguments: first, that morality entails complex, domain-specific rules, concepts, and principles that allow people to make judgments about an infinite number of moral events, much as linguistic grammar allows people to construct infinitely new sentences. Second, there is a ‘poverty of the stimulus’ argument— young children couldn’t possibly have enough time to learn the complexities of human morality through pure learning mechanisms given the early competence they demonstrate—and, therefore, Mikhail argued, at least some of the core attributes of morality are innate, where “innate is used in a dispositional sense to refer to cognitive systems whose essential properties are largely pre-determined by the inherent structure of the mind, but whose ontogenetic development must be triggered and shaped by appropriate experience and can be impeded by unusually hostile learning environments” (2007, p. 144).

Mikhail describes the current dual-process model of moral judgments as inadequate, as models must specify a) which ‘deontic rules’ that pertain to a given moral judgment—deontology and utilitarianism as described here would both qualify as deontic rules according to Mikhail—as well as b) the structural descriptions of the computational operations involved (the semantic relations between means, ends, and

side-effects in each case), and c) the conversion rules that people use to construct those structures in the first place (i.e., how people determine causation, intent, consequences, etc.). Mikhail argues that moral dilemma judgments require lengthy, sophisticated, and abstract representations of complex phenomena (involving various actors, time points, intended effects, side effects, etc.), and are relatively fast, visceral, intuitive and at least partly inaccessible to conscious awareness. Therefore, he proposes that moral dilemma judgments must depend, in part, on innate, domain-specific algorithms. Inconsistent judgments to various dilemma permutations (e.g., personal-impersonal) are resolved by an intuitive, lawyer-like ability to draw “intelligent distinctions between superficially similar cases, although their basis for doing so is often obscure” (p. 150). In other words, Mikhail disagrees with Greene’s dual-process model, arguing that the difference between Greene’s personal and impersonal dilemmas boils down to superficial differences. Moreover, Mikhail takes exception to Greene and colleagues’ contention that deontology and utilitarian judgments stem from the application of existing, domain general processes, rather than specific modules.

Like Nichols and Mallon (2006), Mikhail is quite right to point out the dual-process model fails to carefully identify the exact features of the context that entail a moral violation, nor how those features are represented cognitively. Indeed, dilemma researchers have admitted that if participants represent moral dilemmas differently than researchers envision them, then participants are violating the *closed world assumptions* (Bennis, Medin, & Bartels, 2010) of the dilemma and their answer becomes uninterpretable in terms of the utilitarian/deontological distinction.

That said, Mikhail is perhaps too quick to dismiss Greene and colleague’s distinction between personal and impersonal dilemmas as relevant to the structure of morality. If both deontological and utilitarian judgments stem from essentially identical moral intuitions, Mikhail must account for the growing amount of evidence regarding distinct roles for visceral, affect-laden processes driving one kind of judgment and deliberative processes driving another; he does not. Moreover, Mikhail may be overstating the case when he dismisses the likelihood that humans and chimpanzees

share similar affective reactions to similar circumstances; these reactions may be heuristic in nature (e.g., triggered by the sight or thought of blood) rather than necessarily the product of long, complex chains of computational logic. Therefore, it is conservative to posit the same mechanism to underlie objectively similar displays of emotion (e.g., grimace, smile) in different species (Tomasello, 1991)

Indeed, Chomsky has been criticized as excessively cognitive and computational—it may not be necessary to posit that young children are computing abstract linguistic elements algebraically. There is an alternative perspective that holds that a key human evolutionary adaptation was a single breakthrough—the ability to engage in joint attention with conspecifics, developing a theory of mind, which allows for symbolic representation, and therefore, language, which then becomes the gateway to developing logical thought. According to this perspective, the poverty of the stimulus argument is overstated. Tomasello (2006) summed up this perspective thus:

“Usage-based theories hold that the essence of language is its symbolic dimension, with grammar being derivative. The ability to communicate with conspecifics symbolically (conventionally, intersubjectively) is a species specific biological adaptation. The grammatical dimension of language derives from historical processes of grammaticalization, which create various kinds of grammatical constructions (e.g., the English passive construction, noun phrase construction, or *-ed* past tense construction). As opposed to linguistic rules conceived of as algebraic procedures for combining words and morphemes but that do not themselves contribute to meaning, linguistic constructions are themselves meaningful linguistic symbols – since they are nothing other than the patterns in which meaningful linguistic symbols are used in communication (e.g., the passive construction is used to communicate about an entity to which something happens). In this approach, mature linguistic competence is conceived as a structured inventory of meaningful linguistic constructions - including both the more regular and the more idiomatic

structures in a given language (and all structures in between).

According to the usage-based theory, there is no such thing as universal grammar and so the theoretical problem of how a child links it to a particular language does not exist. It is thus a single process theory of language acquisition, in the sense that children are thought to acquire the more regular and rule-based constructions of a language in the same way they acquire the more arbitrary and idiosyncratic constructions: they learn them.” (Tomasello, 2006, p. ).

Similarly, it may be that people reasoning about morality are not always constructing the elaborate representational schemes that Mikhail describes. Sometimes they may be relying on more basic heuristics—the sight of blood may induce an affective reaction without the need for elaborate reasoning. That said, under those conditions when people are engaged in deliberative reasoning, it seems likely they are sensitive to the various factors outlined by Mikhail in detail that influence moral judgments.

According to this perspective, when people encounter a morally-relevant stimulus, they may experience both a visceral reaction to harm that drives a particular judgment, as well as cognitive reasoning where they construct a careful account of intention, cause, effect, intended outcome, side-effect outcome, and carefully weigh these factors to moderate their judgment. In other words, the dual process model might be compatible with Mikhail’s conception if the complex logical constructions he delineates are conceptualized as occurring at the deliberative level, whereas visceral, affective, or heuristic thinking carries independent variance in moral judgments. Mikhail admits that some of his claims are not conclusive, and that further research is needed to refine understanding of these complex issues. The current work was not designed to test the difference between Greene’s and Mikhail’s theories, but insofar as the conclusions corroborate the dual-process model, they lend credence to Greene’s dual-domain-general-process perspective. As such, they add to the growing debate between these perspectives.

## 1.5 Overview of the Field

Since Greene and colleagues' (2001) landmark paper, a large and growing number of researchers began to investigate the correlates of moral judgments as well as situational factors that influence judgments. Now, over a decade later, moral dilemma research has grown into one of the predominant paradigms for investigating human moral psychology, with a variety of findings providing support for Greene's dual-process model regarding the role of affect and cognition in moral dilemma judgments.

Some moral dilemma research predated Greene and colleagues' (2001) landmark paper. Some early work by Petrinovich and colleagues (O'Neill & Petrinovich, 1998, Petrinovich & O'Neill, 1996, Petrinovich, O'Neill, & Jorgensen, 1993) varied the target of harm in trolley dilemmas, as well as the samples used to assess responses (e.g., American vs. Taiwanese samples). In these studies, people reported that killing one person to save five was most acceptable when that person was socially distant from the self (e.g., stranger, animal), especially if they were a member of a socially abhorrent group (e.g., The Nazi Party). Conversely, people were more reluctant to kill one person to save five when that person was someone similar to themselves (e.g., a kin or a friend). Nonetheless, people were also sensitive to the number of individuals saved, and responses reflected this pattern: when harmful action would save a great number of individuals, participants were most likely to select this option. These findings make sense, given that people tend to view similar others as more deserving of positive treatment than socially distant others (Olson, Hafer, Cheung, & Conway, 2009).

Petrinovich and colleagues argued that their findings were consistent with evolutionary theory: that people would skew their responses in order to maximize inclusive fitness. Their findings were also consistent with the dual-process model: targets closer to the self should elicit more affective responses than targets further from the self, thereby biasing judgments in favor of close targets and against socially distant ones. This pattern also fits with findings that people with damage to emotion centers in the ventromedial prefrontal cortex, whether due to lesions (Ciaramelli et al., 2007; Koenigs et al., 2007) or frontotemporal dementia (Mendez et al., 2005),

make fewer deontological moral judgments, but do not score differently on low-conflict or non-moral dilemmas. People with damage to these brain regions evidence extreme cost-benefit reasoning in real life, as well as a callous disregard for other people's welfare reminiscent of psychopathy (Blair, 1997; Damasio, 1994).

Utilitarian judgments are increased by manipulations that facilitate rational decision-making, either by highlighting the consequences of action (Bartels, 2008, Study 1; Nichols & Mallon, 2006) or by placing decision options side-by-side (Bartels, 2008, Study 3). Conversely, manipulations that impair cognitive operations impair reaction times for utilitarian decisions (Greene et al., 2008) and actually reduce utilitarian judgments (presumably due to decreased utilitarian inclinations, Conway & Gawronski, 2013).

Participants with greater working memory capacity are more likely to make utilitarian decisions than participants with less capacity (Moore et al., 2008), as are participants higher in deliberative as opposed to intuitive thinking styles (Bartels, 2008; Nichols & Mallon, 2006). Participants who endorse utilitarian judgments are more influenced by framing effects, compared to those who endorse deontology, perhaps because the former think harder about the decision (Tanner, Medin, & Iliev, 2008).

Individuals higher in basal testosterone (Carney & Mason, 2010), who are less compassionate (Hermans, Putman, & van Honk, 2006), were also more likely to make utilitarian decisions, as were participants who watched comedy clips, presumably because mirth blunts the negative affect participants experience when considering harm (Valdesolo & Desteno, 2006). Conversely, experiencing positive emotions that facilitate an other-oriented focus, such as *moral elevation*—a warm feeling of awe experienced upon witnessing another's good deeds (Haidt, 2003a)—increases deontological judgments (Strohinger, Lewis, & Meyer, 2011). Anticipating physiological stress also increases deontological judgments, presumably because it makes participants more aware of their visceral reactions to harm, and more likely to sympathize with the victim of harm (Starcke, Ludwig, & Brand, 2012).



Conversely, manipulations that encourage participants to imagine the harmful consequences of action in vivid detail increase deontological responding. Visualizing harm vividly (Amit & Greene, 2012), responding to dilemmas written in an emotionally evocative manner (Bartels, 2008), or responding to dilemmas accompanied by visual depictions of the victim tend to increase deontological judgments—an effect driven by increased deontology, not reduced utilitarianism (Conway & Gawronski, 2013). This is consistent with other work showing stronger negative affect for vividly imagined moral stimuli (Caruso & Gino, 2011). Moreover, when participants made moral judgment under time pressure, they were more likely to make deontological judgments (Suter & Hertwig, 2011), and participants who experienced stronger autonomic arousal also made more deontological judgments (Navarette, McDonald, Mott, & Asher, 2011).

Not all studies have supported a cognitive-affective distinction between utilitarian and deontological judgments. Borg, Hynes, Van Horn, Grafton, and Sinnott-Armstrong (2006) found activation in brain regions related to both emotion and cognition when participants made deontological or utilitarian judgments, and Manfrinati, Lotto, Sarlo, Palomba and Rumiati (2013) found no difference in reported emotional valence or arousal when participants made deontological vs. utilitarian decisions—but these findings should be interpreted with caution, as the way they present these choices is an either-or scenario, rendering them vulnerable to the non-independence error (see below). Therefore, even if they found evidence of emotional responses when people selected one or the other option, it does not mean people felt that emotion *because* they selected that option—it could be that they selected that option *despite* feeling emotion. Moreover, Manfrinati and colleagues presented only a single study, which was underpowered ( $N = 36$ ), so caution should be employed interpreting this null finding.

Overall, a wealth of findings using a variety of techniques—ranging from fMRI to contextual manipulations to individual differences—provide support for Greene and colleagues' (2001) distinction between visceral, affect-laden reactions to harm, and

deliberative comparisons of outcomes. However, it turns out there is a conceptual problem in these studies that undermines how informative previous work can be.

## 1.6 A Problem: The Non-independence Error

A serious problem in traditional moral dilemma studies boils down to an issue of operationalization. When participants indicate that harm to save others is acceptable, researchers have surmised that this judgment arose due to strong utilitarian inclinations that overpowered weaker deontological inclinations. Conversely, when participants indicate that harm to save others is not acceptable, researchers have surmised that this judgment arose due strong deontological inclinations that overpowered weaker utilitarian inclinations. Thus, judgments have been treated as pure measures of process: if a person makes 70% deontological judgments and 30% utilitarian judgments, then researchers would give them a score of .3 *on utilitarianism*, even though they could also be coded as .7 *on deontology* (e.g., Koenigs et al., 2007). This coding implies an inverse relation between the two moral inclinations, in that stronger deontological inclinations imply weaker utilitarian inclinations, and vice versa. Note that when responding to dilemmas, selecting the deontological option (*harm is unacceptable*) also entails also rejecting the utilitarian option (*harm is acceptable*).

Therein lies the rub: for any set of dilemma judgments, researchers cannot tell if the judgments arose from the *strength* of one inclination or the *weakness* of another. Should 30% utilitarian judgments be conceptualized as evidence of strong deontology or as evidence of weak utilitarianism? Note that the problem is not ameliorated by expanding the size of the scale, as some researchers have done (e.g., Bartels & Pizarro, 2011)—scales anchored by *harm is acceptable* at one end and *harm is unacceptable* at the other still treat them as inverse of one another. Granted, it might be possible that the constructs driving deontological and utilitarian judgments *are* the inverse of one another, in which case the current dilemmas' operationalization is sufficient. However, this possibly cannot be tested empirically unless one uses measures that assess the two inclinations independently.

Moreover, Greene and colleagues' (2001) dual-process process theory explicitly argues that deontological judgments are driven by visceral, affect-laden processes, whereas utilitarian judgments are driven by deliberative processes—this was the very insight that served to reconcile the rift between intuition-based and rationalist models of moral psychology. If, indeed, deontology and utilitarian judgments are opposites, then researchers cannot rule out whether they are actually the product of a single process (cf. Kruglanski & Gigerenzer, 2011). Therefore, the way moral dilemma researchers have been measuring deontology and utilitarianism renders them vulnerable to criticisms from either pure intuitionist or pure rationalist camps—and then the field would be right back where it was in 2001, with no clear model integrating the roles of visceral, affect-laden processes with deliberative ones.

Even if Greene's dual-process model is correct, then measuring deontology and utilitarianism as inverse would only be accurate in circumstances where a given person experiences one inclination much more strongly than the other. Instead, consider a person who experiences *both* strong deontological inclinations and strong utilitarian inclinations. Ultimately, despite caring deeply about the fate of five people, she also abhors harming one person so much that she selects *harm is unacceptable*. This person would be coded *low on utilitarianism*—a misclassification. Consider another person who experiences weak deontological *and* weak utilitarian inclinations—they don't particularly care about others' outcomes or about causing harm. Yet, she must choose a response, so she arbitrarily selects the response on the right: *harm is acceptable*. This person would be coded *high on utilitarianism*—another misclassification.

The second example above is problematic if people *who don't mind causing harm* are being labeled *high on utilitarianism*. Indeed, there is a large and growing set of findings suggesting a link between utilitarian moral judgments and psychopathy (Bartels & Pizarro, 2011) and damage to the ventromedial prefrontal cortex (VMPFC) leading to disordered behavior described as *acquired sociopathy* (Koenigs et al., 2011). It is difficult to know what conclusion to draw from this work. One possibility is that psychopathy provides clarity and insight that increases a desire for utilitarian

outcomes. Another possibility is that psychopaths are not especially gifted in terms of deliberative reasoning; rather they lack the emotional responsiveness theorized to motivate aversion to harm.

It seems likely that the second case is true: there is evidence that psychopaths and people with damage to the VMPFC make poor interpersonal decisions despite intact reasoning abilities (Cleckley, 1955; Moll et al, 2006; Saver & Damasio, 1991).

Children with psychopathic tendencies are less likely to distinguish the moral from conventional domains (Blair, 1997), perhaps because they lack the visceral reactions others experience regarding moral vs. non-moral judgments (Arsenio & Ford, 1985). Indeed, there is debate over whether psychopaths are even capable of performing as moral agents, given their emotional impairments (e.g., Glannon, 2008; Ishtiyaque, 2010).

Nonetheless, current methods cannot rule out the possibility that psychopaths experience stronger utilitarian inclinations than other groups, because utilitarianism and deontology are confounded in the measures. Note that these possibilities matter not only with regard to the clarity of researchers' understanding regarding moral psychology; they also convey implications regarding the relation of society to known psychopaths. If the first conclusion is correct—psychopaths are adept at deliberation—then perhaps there is something the rest of society might learn from them with regards to moral decision-making. Conversely, if the second conclusion is correct—psychopaths are deficient in empathic concern—then perhaps society need not consider taking moral advice from psychopaths.

Given the above limitations to traditional moral dilemma measurement, my colleague Bertram Gawronski and I resolved to develop a new measure of moral dilemma responses that was capable of overcoming the non-independence error by independently measuring the strength of deontological and utilitarian inclinations: process dissociation (Jacoby, 1991). If successful, such an approach would provide the advantages of (a) empirically testing whether deontology and utilitarian inclinations are inversely related, unrelated, or even positively related, (b) clarifying

the unique relations between each parameter to determine whether they correspond to constructs described by Greene and colleagues' dual-process model, and (c) clarifying relations between moral judgments and third variables (e.g., psychopathy). In so doing, we hoped to inform the ongoing debate regarding the nature of human moral psychology (for a discussion, see Paxton & Greene, 2011).

## 1.7 Process Dissociation as a Potential Solution

The problems inherent in the traditional approach can be overcome by employing Jacoby's (1991) PD procedure to independently quantify the strength of deontological and utilitarian inclinations within individuals. Jacoby originally developed the process dissociation procedure in order to tease apart the distinct contributions of recollection- versus familiarity-based memory performance. During the 1970s and 1980s, there was a great deal of interest in dissociations between intentional recall and familiarity-based responses on memory tasks (for a review, see Hollender, 1986). Memory performance under impaired processing conditions (e.g., cognitive load or amnesia) was said to be driven wholly by familiarity, whereas unburdened performance was said to be driven wholly by conscious recollection. In other words, researchers regarded tasks as *process pure*. Jacoby (1991) criticized this conception as simplistic because it is unlikely that cognitive load completely eliminates conscious processing, or that automatic processes are silent when participants are asked to make controlled judgments.

Jacoby presented participants with word lists, and then tested their ability to discriminate previously encountered words from new words. Even under cognitive load, participants were more accurate when performance was facilitated by both familiarity and recollection (i.e., they were instructed to include words they consciously recognized as old) than when familiarity was pitted against recollection (i.e., they were instructed to *exclude* words they consciously recognized as old). This indicates that it was a mistake to treat some tasks as pure measures of automatic processing and others as pure measures of controlled processing, because tasks were not process pure.

Jacoby found that participants were capable of correctly identifying more previously-encountered words when familiarity and conscious recollection operated congruently than when one process was pitted against the other. Based on this result, Jacoby surmised that the role of conscious recollection could be ascertained by subtracting the probability of correctly remembering a word when familiarity and recollection conflicted from the probability of correctly remembering a word when familiarity and recollection operated in concert. Once controlled processing has been estimated, automatic processing can be derived by taking the probability of correctly remembering when processes conflict and dividing by the inverse of conscious processing.

Although originally developed to examine memory processing, process dissociation is content agnostic and may be applied to any domain where traditional methods conflate the measurement of two psychological processes. For example, PD has been employed to delineate the partially-overlapping contributions of weapon identification and racial bias when participants indicated whether objects held by White or Black people were guns or cell phones (Payne, 2001). Payne demonstrated that the identification of target objects is an effortful, resource-dependent process, whereas racial bias is unaffected by momentarily available resources. PD has also helped distinguish the role of stereotype category information versus conscious recollection in memory for stereotypic traits (Hense, Penner, & Nelson, 1995), the role of heuristics versus rational calculation in economic decision-making (Ferreira, Garcia-Marques, Sherman, & Sherman, 2006), memory for primary and secondary emotions attributed to in-group and out-group members (Gaunt, Leyens, & Demoulin, 2002), and many other processes (for reviews, see Kelley & Jacoby, 2000; Payne & Bishara, 2009; Yonelinas, 2002). Yet, as far as we are aware, process dissociation has never been applied to moral dilemma responses—a perfect candidate for PD, because responses are theorized to result from the operation of two separate processes, and these processes are conflated in traditional dilemma measures.

Whereas memory researchers conflated memory judgments with processes, morality researchers conflate moral judgments with processes. Moral dilemma response

options are typically dichotomous; participants must select either the deontological or utilitarian response (e.g., Greene et al., 2001). Responses are then treated as process pure: Deontological decisions are presumed to stem from deontological processing; utilitarian decisions from utilitarian processing. On the face of it, this assumption seems reasonable—after all, the dilemmas are designed so that deontology and utilitarianism lead to opposing responses. Yet, this assumption requires inverse relations between deontology and utilitarianism, such that stronger deontological processing is redundant with weaker utilitarian processing. This seems unlikely. If theorists are correct that deontology and utilitarianism reflect affective and cognitive processing, respectively (see Greene, 2007), then these constructs ought to be orthogonal, or even positively related. Therefore, *both* may independently contribute to each dilemma response. Note that approaches that place deontology and utilitarianism on a continuum are still vulnerable to the same issue (e.g., Bartels & Pizarro, 2011).

In order for participants to arrive at a deontological judgment, they must engage in (a) some degree of deontological processing, as well as (b) a lesser degree of utilitarian processing. Likewise, utilitarian judgments require (a) some degree of utilitarian processing, coupled with (b) a lesser degree of deontological processing. Thus, the judgments participants make tell us little about the processing behind those judgments: Deontological judgments may reflect either strong deontological processing or weak utilitarian processing, and utilitarian judgments may reflect either of the opposite cases. Traditional methods of analysis cannot distinguish between these possibilities because deontology is not measured independently of utilitarianism. Determining the strength of deontological and utilitarian processing behind moral dilemma judgments requires a fundamentally different kind of analysis.

The key to process dissociation is employing both *incongruent* trials where the underlying processes lead to divergent responses, as well as *congruent* trials where they lead to the same response. This is a fundamental departure from Greene's personal-impersonal distinction. Both personal and impersonal dilemmas involve a situation where sacrificing one person will save five. This always pits deontological

ethical considerations against utilitarian ones; the difference is that one type is more emotionally evocative, essentially. Compare that to the incongruent-congruent distinction. Incongruent dilemmas pit deontological ethical considerations against utilitarian ones, in that causing harm will lead to greater positive outcomes. Consider the incongruent version of the torture dilemma:

“You are a police officer, and have recently caught a criminal you have been hunting for some time. He is allegedly responsible for rigging a series of explosive devices: some that have already gone off and some that have yet to detonate. He places explosives outside city cafes and sets them to go off at a time when people are drinking coffee on the patios. In this manner, he has injured many people and might injure many more. Now that the criminal is in custody, you want to know where the unexploded bombs are so you can defuse them. He refuses to talk, so you decide to use “aggressive interrogation techniques” like holding his head under water and beating him. Is it appropriate for you to use “aggressive interrogation techniques” in order to find and defuse the unexploded bombs?” (Conway & Gawronski, 2013).

Incongruent dilemmas correspond to Greene and colleagues’ (2001) personal dilemmas and Koenig et al.’s (2007) high-conflict dilemmas. Thus, responses to these dilemmas may be summed and analyzed the traditional way: how many times did people say harm was acceptable or not acceptable when it leads to positive outcomes?

In contrast, congruent dilemmas do not present a conflict between the underlying tendencies: deontological and utilitarian ethical positions agree. They are matched as closely as possible in content and wording to incongruent dilemmas, except that the consequences of harm are far less beneficial. As a result, it is possible to reject harm on either utilitarian or deontological grounds (i.e., the processes that drive deontological and utilitarian judgments will both motivate a judgment that harm is unacceptable). For example, consider the congruent version of the torture dilemma:

“You are a police officer, and have recently caught a criminal you have been hunting for some time. He is allegedly responsible for rigging a series of



explosive devices: some that have already gone off and some that have yet to detonate. He places explosives outside city cafes and sets them to go off at a time when no one is around. His explosives are inside paint cans so that they spray nearby objects with paint. In this manner, he has sprayed many cafes with paint and might spray many more. Now that the criminal is in custody, you want to know where the unexploded bombs are so you can defuse them. He refuses to talk, so you decide to use “aggressive interrogation techniques” like holding his head under water and beating him. Is it appropriate for you to use “aggressive interrogation techniques” in order to find and defuse the unexploded bombs?” (Conway & Gawronski, 2013).

Participants’ judgments in congruent and incongruent moral dilemmas can be illustrated by means of a processing tree (see Figure 1). Each path from left to right depicts judgment outcomes on the two kinds of dilemmas as a function of distinct processes. The three paths in the figure capture the three cases that (1) utilitarianism ultimately drives the response (top path), (2) deontology ultimately drives the response (middle path), and (3) neither utilitarianism nor deontology drives the response (bottom path). *U* depicts the case that utilitarianism drives the response, and *D* depicts the case that deontology drives the response. Conversely, *I – U* depicts the case that utilitarianism does not drive the response, and *I – D* depicts the case that deontology does not drive the response. Using the table on the right side of the figure, it is then possible to use these cases to identify their judgment outcomes for congruent and incongruent dilemmas. In congruent dilemmas, for example, participants will judge harm as unacceptable when utilitarianism drives the response (*U*).

Alternatively, if utilitarianism does not drive the response (*I – U*), harm will still be judged as unacceptable when deontology drives the response (*D*). Harm will be judged as acceptable in congruent dilemmas only when neither utilitarianism (*I – U*) nor deontology (*I – D*) drives the response. Similarly, in incongruent dilemmas, participants will judge harm as unacceptable when utilitarianism does not drive the response (*I – U*) and, at the same time, deontology does drive the response (*D*). However, harm will be judged as acceptable either when utilitarianism drives the

response ( $U$ ), or alternatively when neither utilitarianism ( $1 - U$ ) nor deontology ( $1 - D$ ) drives the response.

By means of the processing paths depicted in Figure 1, it is now possible to create mathematical equations that delineate the probability of a particular overt judgment in congruent and incongruent dilemmas as a function of the two underlying inclinations. For example, the probability of overtly judging harm as unacceptable in a congruent dilemma is represented by the cases where (1) utilitarianism drives the response, and (2) deontology drives the response when utilitarianism fails to drive the response. In algebraic terms, this probability may be represented as:

$$(1) p(\text{unacceptable} \mid \text{congruent}) = U + [(1 - U) \times D]$$

Conversely, the probability of judging harm as acceptable in a congruent dilemma is represented by the case that neither utilitarianism nor deontology drives the response, which can be represented algebraically as:

$$(2) p(\text{acceptable} \mid \text{congruent}) = (1 - U) \times (1 - D)$$

The same logic can be applied to incongruent dilemmas. For example, the probability of judging harm as unacceptable in an incongruent dilemma is represented by the case that deontology drives the response when utilitarianism does not drive the response. Algebraically, this likelihood is represented by the equation:

$$(3) p(\text{unacceptable} \mid \text{incongruent}) = (1 - U) \times D$$

Conversely, the probability of judging harm as acceptable in an incongruent dilemma is represented by the cases that (1) utilitarianism drives the response, and (2) neither deontology nor utilitarianism drives the response. In algebraic terms, this probability is represented as:

$$(4) p(\text{acceptable} \mid \text{incongruent}) = U + [(1 - U) \times (1 - D)]$$

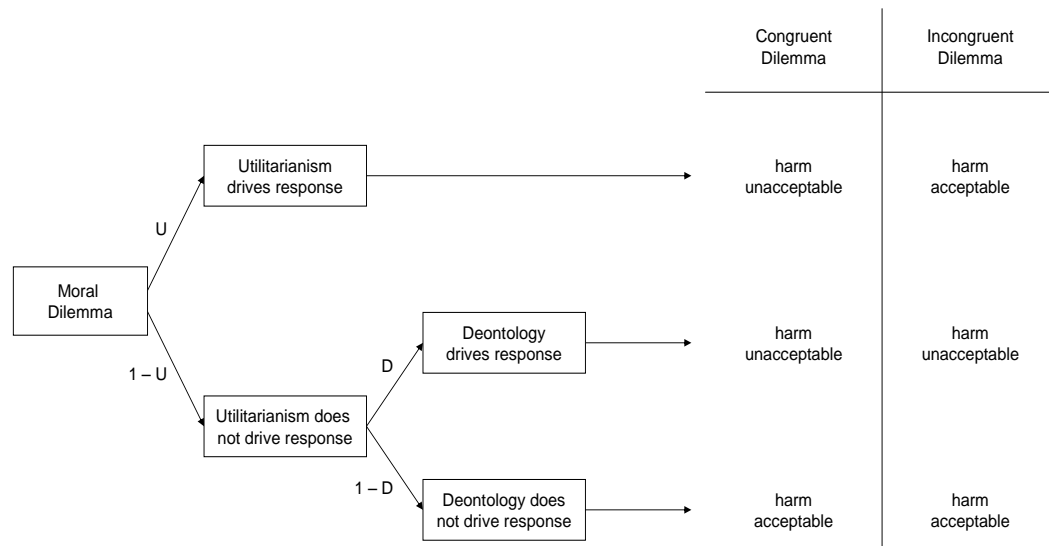


Figure 1: Processing tree illustrating the underlying components leading to judgments that harmful action is either acceptable or unacceptable in congruent and incongruent moral dilemmas. The paths from left to right depict the three cases that (1) utilitarianism ultimately drives the response, (2) deontology ultimately drives the response, and (3) neither utilitarianism nor deontology drives the response.

Using the empirically observed probabilities of participants' *acceptable* and *unacceptable* responses on congruent and incongruent dilemmas, these equations can be used to calculate numerical estimates for the two kinds of moral inclinations by solving algebraically for the two parameters representing deontology (*D*) and utilitarianism (*U*). Specifically, by integrating Equation 3 into Equation 1, the latter can be solved for *U*, leading to the following formula:

$$(5) U = p(\text{unacceptable} \mid \text{congruent}) - p(\text{unacceptable} \mid \text{incongruent})$$

Moreover, by including the calculated value for *U* in Equation 3, this equation can be solved for *D*, leading to the following formula:

$$(6) D = p(\text{unacceptable} \mid \text{incongruent}) / (1 - U)$$

These two formulas provide researchers with a means to quantify the strength of deontological and utilitarian inclinations within participants. For example, if a participant shows an unacceptable response on 7 out of 10 congruent dilemmas (i.e., probability of .70) and on 2 out of 10 incongruent dilemmas (i.e., probability of .20), the above equations would estimate this participant's utilitarian inclination with a value of .50 and his or her deontological inclination with a value of .40 (for a discussion of the metric of each score, as well as other technical details of PD, see Appendix 2B). Such parameter estimates can be calculated for each participant in a given sample, allowing researchers to use them as measurement scores in experimental or individual difference designs. Critically, these scores need not be negatively correlated (i.e., stronger inclinations of one kind are associated with weaker inclinations of the other kind), as implied by the traditional bipolar treatment of moral dilemma responses. Instead, they may vary independently, so that the two parameters may demonstrate unique relations with other variables and distinct effects of experimental manipulations.

It is important to note that the processing paths depicted in Figure 1 represent conditional probabilities that (1) utilitarianism ultimately drives the response, (2) deontology ultimately drives the response, and (3) neither utilitarianism nor

deontology drives the response. The definition of the processing paths in terms of which inclination ultimately drives the response is the reason why there is no fourth path in which both utilitarianism and deontology drive the response. After all, only one of the two moral inclinations can ultimately drive the response. The strength of underlying moral inclinations is estimated by comparing responses on congruent and incongruent dilemmas over multiple trials, allowing researchers to calculate two independent inclination scores on the basis of observed conditional probabilities using the equations outlined in the main text of this article.

Note that this model is agnostic regarding which process comes first temporally. Although, it first delineates the U parameter, and then uses that parameter to calculate the D parameter, this does not imply that utilitarian processing occurs first, and deontological processing coming in later. Indeed, the dual process model suggests that visceral, affective processes associated with deontology occur rapidly and automatically upon noticing a potential moral violation, whereas deliberative processes associated with utilitarianism require additional processing time. The model in Figure 1 is consistent with this perspective.

Yet, the processing paths do imply that utilitarianism will likely drive responses when sufficiently strong; deontology may only drive responses when utilitarianism is relatively weak. Indeed, any application of PD requires a decision as to whether one or the other process dominates responses. In the processing tree depicted in Figure 2, utilitarianism is assumed to dominate, such that deontology may drive the response only if utilitarianism fails to drive the response (U-dominant model). However, it is also possible to construct a PD model in which deontology dominates responses, such that utilitarianism may drive the response only if deontology fails to drive the response (D-dominant model). With regard to the implied outcomes in the table on the right side of the figure, the two PD variants have the same implications for the paths in which either utilitarianism or deontology ultimately drive the response. However, the two variants differ with regard to predicted outcomes when neither utilitarianism nor deontology drives the response. Specifically, the structure of PD implies that, when neither process drives the response, the outcomes are opposite to

those when the subordinate process drives the response. Thus, whereas the U-dominant model implies acceptance of harm in both congruent and incongruent dilemmas (see Jacoby, 1991), the D-dominant model implies acceptance of harm in congruent dilemmas but rejection of harm in incongruent dilemmas (see Lindsay & Jacoby, 1994). Although these differences lead to somewhat different equations for the two parameters (see Payne & Bishara, 2009), the two models produced identical results in the studies reported in this these (with the exception that the two PD parameters of the D-dominant model evince moderate positive correlations across all three studies).

We believe that the U-dominant model reported in the current paper is preferable for two reasons. First, PD models that are structurally equivalent to the U-dominant model have been validated and applied to a wide range of different tasks (e.g., recognition memory, sequential priming, heuristic judgment; for a review, see Payne & Bishara, 2009), whereas PD models that are structurally equivalent to the D-dominant model have been used in only one publication on Stroop performance (Lindsay & Jacoby, 1994). Second, and more importantly, the D-dominant model makes the theoretically implausible assumption that, when neither utilitarianism nor deontology drives responses, participants accept harm in congruent dilemmas, but reject harm in incongruent dilemmas. In other words, the absence of any moral concern would lead to acceptance of major harm but rejection of minor harm. Conversely, the U-dominant model makes the more plausible assumption that participants accept harm in both congruent and incongruent dilemmas when neither moral inclination drives the response. Unconditional acceptance of harm plausibly reflects the absence of moral concern, in that people simply do not care about the harm their actions are causing. Thus, we endorse the U-dominant model for the application of PD to moral dilemmas, especially considering that the U-dominant and D-dominant PD models produced almost identical results in the current work.

## 1.8 The Current Project

Although process dissociation allows for the calculation of independent parameters, it does not guarantee that these parameters, in and of themselves, map onto theorized

constructs. Rather, they must be validated as one would validate any new measure: by examining the unique relations each measured variable shares with other established constructs, as well as examining the impact of theoretically-meaning contextual manipulations. By empirically examining these factors it becomes possible to draw conclusions as to whether the parameters appear to tap constructs expected given the dual-process model (Greene et al., 2001), and, if so, whether they corroborate this model or cast doubt on whether it accurately represents human moral psychology. Once the construct validity of each parameter is established, it becomes possible to employ them to examine theoretical issues in the field at a higher degree of resolution than was possible before.

This leads us to the current project: in five studies we (myself and two sets of co-authors) developed, validated, and applied a process dissociation measure of the deontological and utilitarian inclinations underpinning judgments on moral dilemmas. In Conway and Gawronski (2013) Study 1, we examined the relations between the PD parameters and theoretically relevant individual-difference variables (e.g., the utilitarian parameter uniquely predicted cognitive load, whereas the deontology parameter uniquely predicted empathic concern and perspective-taking). In Study 2, we determined whether a cognitive load manipulation would selectively reduce utilitarian inclinations, whereas in Study 3, we examined whether a manipulation that increased the vividness of harm selectively increases the deontology parameter. If successful, these findings would suggest that the deontology parameter is successfully tapping affective reactions to harm, whereas the utilitarian parameter is tapping cognitive evaluations of outcomes (consistent with the dual-process model, Green et al., 2001). Next, we aimed to apply this technique to clarify the relations between utilitarianism and deontology with both pro- and anti-social tendencies. In Study 1 of Conway, Bartels, and Pizarro (under review), participants scoring higher in Machiavellianism, psychopathy, and meaninglessness also made more overt utilitarian judgments (replicating past findings), but process dissociation revealed that this relation was due to *decreased deontology* rather than *increased utilitarianism* among people high in antisocial personality traits. Study 2 demonstrated that the

deontology and utilitarian parameters each correlated with different kinds of prosociality.



## References

- Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine de Gruyter.
- Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, 23, 861-868. doi:10.1177/0956797611434965
- Aristotle (1962). *Nichomachean ethics*. (Martin Oswald, Trans.) Indianapolis, IN: Bobbs-Merrill. (Original work published 4th Century B.C.).
- Arsenio, W. F. & Ford, M. E. (1985). The role of affective information in social-cognitive development: Children's differentiation of moral and conventional events. *Merrill-Palmer Quarterly*, 31, 1-17.
- Aquinas, T. (1947). *Summa theologiae*. New York: Benzinger Brothers (Originally published in 1265-1272).
- Aquino, K., & Reed, A. II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83, 1423-1440. doi:10.1037//0022-3514.83.6.1423
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390-1396.
- Bargh, J. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In J. R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition*, 2nd edition. (pp. 1-40). Hillsdale, NJ: Erlbaum.
- Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology*, 32, 104-128.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70, 1-16.
- Baron, J. (2000). *Thinking and deciding*. Cambridge, United Kingdom: Cambridge University Press.
- Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, 108, 381-417. doi:10.1016/j.cognition.2008.03.001

- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, 121, 154-161. doi:10.1016/j.cognition.2011.05.010
- Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological bulletin*, 115(2), 243-267.
- Bazerman, M. H., & Greene, J. (2010). In favor of clear thinking: Incorporating moral rules into a wise cost-benefit analysis. Commentary on Bennis, Medin, & Bartels (2010). *Perspectives on Psychological Science*, 5, 209-212. doi:10.1177/1745691610362362
- Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, 5, 187-202. doi:10.1177/1745691610362354
- Bergman, R. (2002). Why be moral? A conceptual model from developmental psychology. *Human development*, 45, 104-124.
- Blair, R. J. R. (1997). Moral reasoning and the child with psychopathic tendencies. *Personality & Individual Differences*, 22, 731-739.
- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, 88, 1-45.
- Blasi, A. (1995). Moral understanding and the moral personality: The process of moral integration. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Moral development: An introduction*. Boston, MA: Allyn and Bacon.
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, 18, 803-817. doi:10.1162/jocn.2006.18.5.803
- Burnstein, E., Crandall, C. & Kitayama, S. (1994). Some neo-darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of personality and social psychology*, 67(5), 773-789.
- Cameron, C. D. & Payne, D. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, 100, 1-15. doi:10.1037/a0021643
- Carney, D. R., & Mason, M. F. (2010). Moral decisions and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology*, 46, 668-671. doi:10.1016/j.jesp.2010.02.003

- Côté, S., Piff, P., & Willer, R. (2013). For whom do the ends justify the means? Social class and utilitarian moral judgment. *Journal of Personality and Social Psychology*, 104, 490-503.
- Colby, A. & Damon, W. (1992). *Some do care: Contemporary lives of moral commitment*. New York: The Free Press.
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, 104, 216-235. doi:10.1037/a0031021
- Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium* (pp. 3-39). Hillsdale, NJ: Erlbaum.
- Chomsky, N. (1964). *Current Issues in Linguistic Theory*. The Netherlands: Mouton & Co.
- Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2, 84-92. doi:10.1093/scan/nsm001
- Cleckley, H. (1955). *The mask of sanity*. St. Louis: C. V. Mosby.
- Cushman, F., Gray, K., Gaffey, A. & Mendes, W. B. (2012). Simulating murder: The aversion to harmful actions. *Emotion*, 12, 2-7.
- Cushman, F., Young, L., & Greene, J. D. (2010). Our multi-system moral psychology: Towards a consensus view. In J. Doris, G. Harman, S. Nichols, J. Prinz, W. Sinnott-Armstrong, & S. Stich (Eds.), *The Oxford handbook of moral psychology* (pp. 47-71). Oxford: Oxford University Press.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, 17, 1082-1089. doi:10.1111/j.1467-9280.2006.01834.
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behaviour caused by frontal lobe damage fail to respond automatically to social stimuli. *Behavioural brain research*, 41, 81-94.
- Darwin, C. (1874/1998). *The descent of man*. Amherst, NY: Prometheus Books.
- Dawkins, R. (1989). *The selfish gene*. Oxford: Oxford University Press.

- de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, Mass: Harvard University Press.
- Devine, P. G. (1989). Stereotypes and Prejudice: Their Automatic and Controlled Components. *The Journal of Personality and Social Psychology*, 56, 5-18.
- Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, 91, 797–813.
- Flack, J. C. & de Waal, F. B. M. (2000). ‘Any animal whatever.’ In Katz, L. D. (Ed.) *Evolutionary origins of morality* (p. 1-30). Bowling Green, OH: Imprint Academic.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5-15.
- Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Spota, S., Zago, S., Lucchiari, C., Consonni, D., Nordio, F., Pravettoni, G., Cappa, S. & Priori, A. (2010). Gender related differences in moral judgments. *Cognitive Processing*, 11, 219-226. doi:10.1007/s10339-009-0335-2
- Galotti, K. M. (1989). Approaches to studying formal everyday reasoning. *Psychological Bulletin*, 105, 331-351.
- Gangestad, S. W., Haselton, M. G., & Buss, D. M. (2006). Evolutionary foundations of cultural variation: Evoked culture and mate preferences. *Psychological Inquiry*, 17, 75-95.
- Gaunt, R., Leyens, J., & Demoulin, S. (2002). Intergroup relations and the attribution of emotions: Control over memory for secondary emotions associated with the ingroup and outgroup. *Journal of Experimental Social Psychology*, 38, 508–514.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women’s development*. Cambridge, MA: Harvard University Press.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24(3), 153-172.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women’s development*. Cambridge, MA: Harvard University Press.
- Glannon, W. (2008). Moral responsibility and the psychopath. *Neuroethics*, 1, 158-166.

- Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS ONE*, 8, e60418. doi:10.1371/journal.pone.0060418
- Glenn, A. L., Raine, A., & Schug, R. A. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry*, 14, 5-9.
- Glenn, A.L., Koleva, S., Iyer, R., Graham, J., Ditto, P.H. (2010). Moral identity in psychopathy. *Judgment and Decision Making*, 5, 497-505.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029-1046.
- Gray, K., Young, L., Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101-124.
- Greene, J.D. (2003) From neural "is" to moral "ought": What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, 4, 847-850.
- Greene, J. D. (2007a). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, 11, 322-323. doi:10.1016/j.tics.2007.06.004
- Greene, J. D. (2007b). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*. MIT Press, Cambridge, MA.
- Greene, J. and Haidt, J. (2002) How (and where) does moral judgment work? *Trends in Cognitive Science*, 6, 517-523.
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111, 364-371. doi:10.1016/j.cognition.2009.02.001
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107, 1144-1154. doi:10.1016/j.cognition.2007.11.004
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389-400.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105-2108.

- Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition. *Psychological Review*, 102, 4-27.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834.  
doi: 10.1037//0033-295X. 108.4.814
- Haidt, J. D. (2003a). Elevation and the positive psychology of morality. In C. L. M. Keyes & J. D. Haidt (Eds.), *Flourishing: Positive psychology and the life well lived* (pp. 275–289). Washington DC: American Psychological Association.
- Haidt, J. (2003b). The emotional dog does learn new tricks: A reply to Pizarro and Bloom (2003). *Psychological Review*, 110, 197-198.  
doi:10.1037/0033-295X.110.1.197
- Haidt, J. et al. (2007). The new synthesis in moral psychology. *Science*, 316, 998-1002. doi:10.1126/science.1137651
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98-116.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, Fall, 55–66.
- Haidt, J., & Kesebir, S. (2010). Morality. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.) *Handbook of Social Psychology*, 5<sup>th</sup> Edition. Hoboken, NJ: Wiley. Pp. 797-832.
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals<sup>1</sup>. *Journal of Applied Social Psychology*, 31, 191-221.
- Hardy, S. A., & Carlo, G. (2005). Identity as a source of moral motivation. *Human Development*, 48, 232–256. doi:10.1159/000086859
- Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628.
- Hamilton, W. D. (1964). The evolution of social behavior. *Journal of Theoretical Biology*, 7, 1-52.
- Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. *Personality and Social Psychology Bulletin*, 32(10), 1402-1413.

- Hense, R. L., Penner, L. A., & Nelson, D. L. (1995). Implicit memory for age stereotypes. *Social Cognition*, 13, 399–415.
- Herzog, H. A., & Golden, L. L. (2009). Moral emotions and social activism: the case of animal rights. *Journal of Social Issues*, 65, 485-498.
- Hoffman, M. L. (1982). Development of prosocial motivation: Empathy and guilt. In N. Eisenberg (Ed.), *The development of prosocial behavior*. New York: Academic Press.
- Hofmann, W., & Baumert, A. (2010). Immediate affect as a basis for intuitive moral judgment: An adaptation of the affect misattribution procedure. *Cognition and Emotion*, 24, 522-535. doi:10.1080/02699930902847193
- Horberg, E. J., Oveis, C., & Keltner, D. (2011). Emotions as moral amplifiers: An appraisal tendency to the influence of distinct emotions upon moral judgment. *Emotion Review*, 3, 1–8.
- Hume, D. (1969). *A treatise of human nature*. (Trans.) London: Penguin. (Original work published 1739 & 1740).
- Inbar, Y., Pizarro, D.A., & Bloom (2012). Disgusting smells cause decreased liking of gay men. *Emotion*, 12, 23-27.
- Ishtiyaque, H. (2010). Psychopathy, ethical perception, and moral culpability. *Neuroethics*, 3, 135-150.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513-541.
- Kagan, S. (1998). *Normative Ethics*. Westview Press: Boulder, CO.
- Kameda, T., Takezawa, M., & Hastie, R. (2003). The logic of social sharing: An evolutionary game analysis of adaptive norm development. *Personality and Social Psychology Review*, 7, 2-19.
- Kant, I. (1785/1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill.
- Kelley, C. M., & Jacoby, L. L. (2000). Recollection and familiarity: Process-dissociation. In E. E. Tulving, E. Fergus, & I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 215–228). New York: Oxford University Press.
- Killen, M., & de Waal, F. B. M. (2000). The evolution and development of morality. In F. Aureli, & F. B. M. de Waal (Eds.), *Natural conflict resolution*. (pp. 352-372). Berkeley, CA, US: University of California Press.

- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446, 908-911. doi:10.1038/nature05631
- Kogut, T., & Ritov, I. (2005). The “identified victim” effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, 18, 157–167. doi:10.1002/bdm.492
- Kohlberg, L. (1969). Stage and sequence: The cognitive–developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 347–480). Chicago: Rand McNally.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151-235). New York: Academic Press.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151-235). New York: Academic Press.
- Kohlberg, L. & Candee, D. (1984). The relationship of moral judgment to moral action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Morality, moral behaviour, and moral development*. New York: John Wiley.
- Kohlberg, L., & Hersh, R. H. (1977). Moral Development: A Review of the Theory. *Theory into Practice*, 16: *Moral Development*, 53-59.
- Krebs, D. (2008). Morality: An evolutionary account. *Perspectives on Psychological Science*, 3, 149-172.
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Mahwah, NJ: Erlbaum.
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgments are based on common principles. *Psychological Review*, 118, 97-109. doi: 10.1037/a0020762
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96, 674-689.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Lapsley, D. K. (1996). *Moral psychology*. Boulder, CO: Westview.



- Lee, R. B., & DeVore, I. (Eds.). (1968). *Man the hunter*. Oxford, England: Aldine.
- Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., Blair, R.J., (2006). The neural basis of implicit moral attitude—an IAT study using event-related fMRI. *NeuroImage* 30, 1449–1457.
- Macy, M. W., & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review*, 63(5), 638–660.
- Maner, J. K., Luce, C. L., Neuberg, S. L., Cialdini, R. B., Brown, S., & Sagarin, B. J. (2002). The effects of perspective taking on motivations for helping: Still no evidence for altruism. *Personality and Social Psychology Bulletin*, 28, 1601–1610.
- McAlister, A. L., Bandura, A., & Owen, S. V. (2006). Mechanisms of moral disengagement in support of military force: The impact of Sept. 11. *Journal of Social & Clinical Psychology*, 25, 141–165.
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological bulletin*, 127, 249–266.
- McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, 45, 577–580. doi:10.1016/j.jesp.2009.01.002
- Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology*, 18, 193–197.
- Midgley, M. (1991). The origin of ethics. In: P. Singer. (Ed.) *A companion guide to ethics* (p. 1–13). Oxford: Blackwell Reference.
- Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *TRENDS in Cognitive Sciences*, 11, 143–152. doi:10.1016/j.tics.2006.12.007
- Mill, J. S. (1861/1998). *Utilitarianism*. In R. Crisp (Ed.), New York: Oxford University Press.
- Miller, G.F. (2007). The sexual selection of moral virtues. *Quarterly Review of Biology*, 82, 97–125. Moll, J., de Oliveira-Souza, R. and Eslinger, P. (2003). Morals and the human brain: A working model. *NeuroReport*, 14, 299.
- Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, 19, 549–57. doi:10.1111/j.1467-9280.2008.02122.x

- Mullen, E., & Skitka, L. J. (2006). Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, identification, or affect? *Journal of Personality and Social Psychology*, 90, 629–643. doi:10.1037/0022-3514.90.4.629
- Murphy, S. T. & Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 64, 723-729.
- Navarette, C. D., McDonald, M. M., Mott, M. L. & Asher, B. (2011). Morality: Emotion and action in a simulated three-dimensional “trolley problem.” *Emotion*, Vol. 12, No. 2, pp. 364–370.
- Narvaez, D. (2008). The social intuitionist model and some counter-intuitions. In W. Sinnott Armstrong (Ed.), *Moral psychology—Vol. 2: The cognitive science of morality: Intuition and diversity* (pp. 233–240). Cambridge, MA: MIT Press.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, 84, 221–236. Doi:10.1016/S0010-0277(02)00048-3
- Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100, 530-542. doi:10.1016/j.cognition.2005.07.005
- Olson, J. M., Hafer, C. L., Cheung, I., & Conway, P. (2009). Deservingness, the scope of justice, and actions toward others. In A. C. Kay, D. R. Bobocel, M. P. Zanna, & J. M. Olson (Eds.), *The psychology of justice and legitimacy: The Ontario symposium* (Vol. 11, pp. 125-149). New York: Psychology Press.
- O'Neill, P., & Petrinovich, L. (1998). A preliminary cross-cultural study of moral intuitions. *Evolution and Human Behavior*, 19, 349-367.
- Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes*, 109, 134–141.
- Paxton, J. M., & Greene, J. D., (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Sciences*, 2, 1-17. doi:10.1111/j.1756-8765.2010.01096.x
- Paxton, J. M., Ungar, M., & Greene, J. D. (2011). Reflection and reasoning in moral judgment. *Cognitive Science*, 36, 1–15. doi:10.1111/j.1551-6709.2011.01210.x
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181–192. doi:10.1037//0022-3514.81.2.181

- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89, 277-293.
- Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, 20, 272-314. doi:10.1080/10463280903162177
- Pellizzoni, S., Siegal, M., & Surian, L. (2010). The contact principle and utilitarian moral judgments in young children. *Developmental Science*, 13, 265–270. doi:10.1111/j.1467-7687.2009.00851.x
- Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology & Sociobiology*, 17, 145-171.
- Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, 64, 467-478.
- Piaget, J. (1965). *The moral judgement of the child*. (M. Gabain, Trans.) New York: Free Press. (Original work published 1932).
- Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on Haidt (2001). *Psychological Review*, 110, 193–196. doi:10.1037/0033-295X.110.1.193
- Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington, DC: American Psychological Association.
- Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39, 653–660.
- Plato (1949). *Timaeus*. (Benjamin Jowett, Trans.) Indianapolis: Bobbs-Merrill. (Original work published 4th Cent. BC).
- Prinz, J. (2006). The emotional basis of moral judgment. *Philosophical Explorations*, 9, 29-43.
- Reeder, G. D., Kumar, S., Hesson-McInnis, M. S., & Trafimow, D. (2002). Inferences about the perceived morality of an aggressor: The role of perceived motive. *Journal of personality and social psychology*, 83(4), 789-803.
- Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.

- Rest, J. R. (1979). *Development in Judging Moral Issues*. Minneapolis: University of Minnesota Press.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76, 574-586.
- Russel, P. S., & Giner-Sorolla, R. (2011). Anger is more flexible than disgust. *Social Psychological and Personality Science*, 2, 360-364.  
Doi:10.1177/1948550610391678
- Saver, J. L. & Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, 29, 1241-1249. doi:10.1016/0028-3932(91)90037-9
- Salerno, J. M., & Peter-Hagene, L. C. (2013). The interactive effect of anger and disgust on moral outrage and judgments. *Psychological Science*.  
doi:10.1177/0956797613486988
- Saltzstein, H. D., & Kasachkoff, T. (2004). Haidt's moral intuitionist theory: A psychological and philosophical critique. *Review of General Psychology*, 8, 273-282. doi:10.1037/1089-2680.8.4.273
- Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34, 1096-1109.
- Shweder, R. A. (2002). The nature of morality: The category of bad acts. *Medical Ethics*, 9, 6-7.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and Health* (pp. 119-169). New York: Routledge.
- Sherman, G. D., & Clore, G. L. (2009). The color of sin white and black are perceptual symbols of moral purity and pollution. *Psychological Science*, 20, 1019-1025.
- Skitka, L. J., Bauman, C. W., & Mullen, E. (2008). Morality and justice: An expanded theoretical perspective and review. In K. A. Hedgvedt & J. Clay-Warner (Eds.), *Advances in Group Processes*, 25 (pp. 1 - 27). Bingley, UK: Emerald Group Publishing Limited.
- Skitka, L. J., Bauman, C., W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88, 895-917. doi:10.1037/0022-3514.88.6.895

- Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA, US: Harvard University Press.
- Sousa, P., Holbrook, C., & Piazza, J. (2009). The morality of harm. *Cognition*, 113, 80-92. doi:10.1016/j.cognition.2009.06.015
- Starcke, K., Ludwig, A., & Brand, M. (2012). Anticipatory stress interferes with utilitarian moral judgment. *Judgment and Decision Making*, 7, 61-68. doi:10.1016/j.psyneuen.2010.07.010
- Strohminger, N., Lewis, R. L., & Meyer, D. E. (2011). Divergent effects of different positive emotions on moral judgment. *Cognition*, 119, 295-300. doi:10.1016/j.cognition.2010.12.012
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119, 454-458. doi:10.1016/j.cognition.2011.01.018
- Tanner, C., Medin, D. L., & Iliev, R. (2008). Influence of deontological and consequentialist orientations on act choices and framing effects: When principles are more important than consequences. *European Journal of Social Psychology*, 38, 757-769. doi:10.1002/ejsp.493
- Thomson, J. J. (1986). Killing, letting die, and the trolley problem, in William Parent, (ed.), *Rights, Restitution, and Risk: Essays in Moral Theory* (Cambridge: Harvard University Press).
- Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology*. New York: Wiley.
- Trivers, R. (1971). The evolution of reciprocal altruism. *Review of Biology*, 46, 35-57.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.
- Ugazio, G., Lamm, C., & Singer, T. (2012). The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion*, 12, 579-590.
- Valdesolo, P. & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17, 476-477. doi:10.1111/j.1467-9280.2006.01731.x
- Van den Bos, K. (2003). On the subjective quality of social justice: The role of affect as information in the psychology of justice judgments. *Journal of Personality and Social Psychology*, 85, 482-498.

- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16, 780–784.
- Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature*, 308(5955), 181-184.
- Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.
- Wright, R. (1994). *The moral animal: The new science of evolutionary psychology*. New York, NY, US: Pantheon Books.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.

## Notes

<sup>1</sup> Typically, researchers describe the dual-process model as suggesting that affective reactions to harm drive deontological judgments, whereas cognitive evaluations of outcomes drive utilitarian judgments (e.g., Greene, 2007a)—but this is not to say that affect plays no role in utilitarian judgment or that cognition plays no role in deontological judgments. Greene and colleagues (2004) pointed out that on one level, all psychological experiences qualify as “cognition” in that all psychology involves information processing. Yet, they maintained that a distinction between cognition and emotion is possible despite being poorly understood:

“Throughout this article, we have relied on the familiar distinction between ‘emotion’ or ‘affect’ on the one hand and ‘cognition’ on the other. This distinction has proven useful, and yet it may be somewhat artificial. The term ‘cognition’ is often defined in terms of ‘information processing,’ but all of the processes considered here, including those that we have labeled ‘emotional,’ involve information processing, thus calling into question the usefulness of this definition of ‘cognition.’ Alternatively, one might render the emotion/cognition distinction in terms of a contrast between, on the one hand, representations that have direct motivational force and, on the other hand, representations that have no direct motivational force of their own, but that can be contingently connected to affective/emotional states that do have such force, thus producing behavior that is both flexible and goal directed. According to this view, the emotion/cognition distinction is real, but it is a matter of degree and, at the present time, not very well understood. It is within a framework of this sort that we retain and utilize the emotion/cognition distinction while recognizing that this distinction is far from clear cut” (2004, p. 397-398).

Therefore, Greene and colleagues maintain a distinction between ‘emotion’ and ‘cognition’ by arguing the former construct entails direct motivational force, whereas the latter entails motivation via more distal related processes. Indeed, Greene and

colleagues (2004) argue that the motivational force in utilitarian judgments stems from affect elicited by the products of cognition:

“Like David Hume (Hume, 1978), we suspect that all action, whether driven by ‘cognitive’ judgment or not, must have some affective basis. Even a cold, calculating utilitarian must be independently motivated, first, to engage in the reasoning that utilitarian judgment requires and, second, to respond in accordance with such judgment” (2004, p. 397).

Moreover, although Greene and colleagues do not directly examine the issue, they appear to tacitly accept that cognition plays a role in deontological judgments as well. Nichols and Mallon (2006) developed this argument much further: they argued that deontological intuitions require not only negative emotional responses, but also the appraisal that a moral rule has been violated. Otherwise, any time people experienced negative affect, they would judge that a moral rule has been violated (for example, during ritual circumcision). Hence, Nichols and Mallon described deontological moral rules as *affect-backed rules*, to distinguish them from mere negative affect without perception of a rule violation, and from *non-affect-backed rules*. Nichols and Mallon found that when a moral rule is violated and people experience strong negative affect, they display *all-in impermissibility*, judging harm to be morally wrong regardless of the consequences: a deontological response. When the same rule is violated but people do not experience strong affect, they display *weak impermissibility*, in that they judge harm to be morally wrong, but acceptable, all things considered—a consequentialist response. Thus, their model involves three components: a) the recognition that a violation has occurred, b) an affective reaction, and c) a cognitive response. Nichols and Mallon’s conceptualization of affect-backed moral rules offers a conceptual advance over traditional moral dilemma research. Yet, most moral dilemma research (including the current work) tacitly assumes that participants perceive a moral violation has occurred and focuses on subsequent processing—a reasonable, if untested, assumption. Hence, the current work is fully compatible with Nichols and Mallon’s and Greene and colleagues (2004) insights regarding the tangled roles of cognition and affect in moral judgments. Nonetheless,



following Greene and colleagues (2004), the current work draws a distinction between affective and cognitive processing until such time as a more fleshed-out framework for the interplay between these forces in moral judgment becomes widespread. In the meantime, it may be wise for moral dilemma researchers to more regularly employ Greene's phrase *matter of degree*—for both deontological and utilitarian judgments seem to entail some degree of both cognition and emotion; they appear to differ in the strength of these processes rather than their absolute presence or absence.

<sup>2</sup> Readers familiar with philosophy will note a fundamental discrepancy between the dual-process model of moral judgments and the philosophical positions associated with each process. Kant (1785/1959) argued that the only route to deontological moral judgments was through rational appreciation of duty, whereas Greene and colleagues (2001) argued that deontological decisions arise from affective reactions to harm. Similarly, utilitarian ethical considerations have a historical association with affect, given that utility may be conceptualized as overall happiness or wellbeing (Hume, 1969; Mill, 1861/1998). Greene and colleagues were well aware of this discrepancy—so much so that Greene titled his dissertation publication *The Secret Joke of Kant's Soul* (2007b). In this paper, Greene clarified:

“Because I am interested in exploring the possibility that deontology and consequentialism are psychological natural kinds, I will put aside their conventional philosophical definitions and focus instead on their relevant functional roles. As noted earlier, consequentialists and deontologists have some characteristic practical disagreements. For example, consequentialists typically say that killing one person in order to save several others may be the right thing to do, depending on the situation. Deontologists, in contrast, typically say that it's wrong to kill one person for the benefit of others, that the 'ends don't justify the means.' Because consequentialists and deontologists have these sorts of practical disagreements, we can use these disagreements to define consequentialist and deontological judgments functionally. For the purposes of this discussion, we'll say that

consequentialist judgments are judgments in favor of characteristically consequentialist conclusions (e.g., “Better to save more lives”) and that deontological judgments are judgments in favor of characteristically deontological conclusions (e.g., “It’s wrong despite the benefits”)... [Therefore,] when I refer to something as a ‘deontological judgment’ I am saying that it is a characteristically deontological judgment and am not insisting that the judgment in question necessarily meets the criteria that philosophers would impose for counting that judgment as deontological” (2007b, p. 38-39).

However, given the etymological history of the words deontology and utilitarianism, it is actually rather presumptuous of psychologists and other scientists to use these terms interchangeably with moral dilemma judgments—these terms refer to more than mere judgments in philosophy; they also entail assumptions regarding the process by which people arrive at a given judgment. The problem with conflating philosophical terminology with dilemma judgments is that contentious conclusions may be drawn. Greene and colleagues (2004) argued, “Should this [dual-process] account prove correct, however, it will have the ironic implication that the Kantian, ‘rationalist’ approach to moral philosophy is, psychologically speaking, grounded not in principles of pure practical reason, but in a set of responses that are subsequently rationalized” (2004, p. 398). Greene went on to argue that as a consequence, deontology ought to be discounted as a genuine moral position; rather, people should be encouraged to make utilitarian judgments (Greene, 2003)—a notion that garnered stiff opposition (e.g., Bennis, Medin, & Bartels, 2008).

In the current work there are no such assumptions regarding the legitimacy of one or another ethical position; the focus is purely descriptive, aimed at understanding the factors that influence moral dilemma judgments. To this end, it might be advantageous to adopt more neutral labels, such as *judgments that harm is not acceptable* in place of *deontological judgments*, and *judgments that harm is acceptable* in place of *utilitarian judgments*. Nonetheless, given that the practice of using these terms interchangeably is so common in the field as to be essentially

ubiquitous, for the time being I have co-opted this language despite the complex theoretical baggage it entails. I look forward to future discussion aimed at revising the terminology employed, but recognize that the current work is not an appropriate forum for this discussion.

<sup>3</sup> Greene and colleagues (2009) argued that the doctrine of double effect fails to adequately explain moral judgments, as intention to harm a person as a means to saving the others, rather than as a foreseen but unintended consequence of action, impacted judgments of acceptability only when harm was dealt via personal force rather than through a mechanical intermediary.

<sup>4</sup> Greene and colleagues (2001, 2004) also presented reaction time data suggesting that people took longer on impersonal than personal dilemmas, but later retracted this claim (Greene and colleagues, 2008) after Liane Young pointed out inconsistencies in their dilemma classification system (see also McGuire, Langdon, Coltheart, & Mackenzie, 2009). When they reanalyzed their data, the fMRI and behavioral data stood up, but the reaction time data did not. Therefore, we will not consider these data further.

## Chapter 2

### 2 Deontological and Utilitarian Inclinations in Moral Decision-Making: A Process Dissociation Approach

Paul Conway and Bertram Gawronski, The University of Western Ontario

Copyright © 2013 by the American Psychological Association, Reproduced with Permission. The official citation that should be used in referencing this material is Conway, P., & Gawronski, B. (2013). *Journal of Personality and Social Psychology*, 104, 216–235. doi:10.1037/a0031021. No further reproduction or distribution is permitted without written permission from the American Psychological Association.

Article available at <http://psycnet.apa.org/journals/psp>

#### 2.1 Abstract

Dual-process theories of moral judgment suggest that responses to moral dilemmas are guided by two moral principles: the principle of *deontology* states that the morality of an action depends on the intrinsic nature of the action (e.g., harming others is wrong regardless of its consequences); the principle of *utilitarianism* implies that the morality of an action is determined by its consequences (e.g., harming others is acceptable if it increases the well-being of a greater number of people). Despite the proposed independence of the moral inclinations reflecting these principles, previous work has relied on operationalizations in which stronger inclinations of one kind imply weaker inclinations of the other kind. The current research applied Jacoby's (1991) process dissociation procedure to independently quantify the strength of deontological and utilitarian inclinations within individuals. Study 1 confirmed the usefulness of process dissociation for capturing individual differences in deontological and utilitarian inclinations, revealing positive correlations of both inclinations to moral identity. Moreover, deontological inclinations were uniquely related to empathic concern, perspective-taking, and religiosity, whereas utilitarian inclinations were uniquely related to need for cognition. Study 2 demonstrated that

cognitive load selectively reduced utilitarian inclinations, with deontological inclinations being unaffected. In Study 3, a manipulation designed to enhance empathy increased deontological inclinations, with utilitarian inclinations being unaffected. These findings provide evidence for the independent contributions of deontological and utilitarian inclinations to moral judgments, resolving many theoretical ambiguities implied by previous research.

## Chapter 3

### 3 Are Psychopaths Really More Utilitarian? Process Dissociation Clarifies the Relations Between Utilitarianism, Deontology, Anti-Social Personality Traits, and Prosociality

Paul Conway, Department of Psychology, The University of Western Ontario, Canada

Daniel M. Bartels, Graduate School of Business, Columbia University, United States

David A. Pizarro, Department of Psychology, Cornell University, United States

#### 3.1 Abstract

Recent research using traditional sacrificial moral dilemmas to measure of moral judgment has demonstrated a puzzling finding—that people who score higher in antisocial personality traits (such as psychopathy) also make more utilitarian decisions. To shed light on this finding, we conducted two studies using a process dissociation technique which allowed us to independently estimate whether utilitarian moral judgments are driven by increased utilitarian concerns or as a result of reduced deontological inclinations. Study 1 demonstrated that the relation between utilitarian decisions and antisocial traits was not driven by *increased* utilitarian inclinations, but rather *decreased* deontological ones. Study 2 demonstrated that utilitarian and deontological inclinations each predicted different kinds of prosocial motivation. Together, these results suggest that utilitarian judgments among participants with antisocial personality traits arise not as a result of a commitment to maximizing the overall outcome, but rather due to a blunted aversion to causing harm.

#### 3.2 Introduction

Consequentialist ethical positions such as *utilitarianism* (e.g., Mill, 1861/1998) impose the moral requirement that people act in ways that maximize overall outcomes. Thus, utilitarianism mandates killing one person to save several others. Likewise, it mandates forgoing one's disposable income to help the less fortunate. Conversely, rule-based ethical positions such as *deontology* (e.g., Kant, 1785/1959) impose the moral

requirement that people abide by a moral rule (such as do no harm) regardless of whether doing so leads to the best overall consequences. According to deontology, killing an innocent person is forbidden—even to save multiple lives—but there is no mandate to forgo one’s disposable income to benefit others (Kagan, 1998).

Recently, moral psychologists have adduced evidence suggesting that deontological and utilitarian ethical judgments arise from differing psychological processes: *Deontological* judgments are associated with fast, visceral, affective processing, whereas *utilitarian* judgments are associated with slow, effortful, and resource-dependent processing (see, e.g., Conway & Gawronski, 2013; Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Greene et al., 2001, 2007; Koenigs et al., 2007; Mendez, Anderson, & Shapria, 2005; cf. Baron, 2011; Kahane et al., 2012). Researchers have collected this evidence through the use of *high-conflict* moral dilemmas, designed to pit one ethical position against the other (e.g., Koenigs et al., 2007). High-conflict dilemmas entail dealing harm in order to achieve a greater positive outcome; participants indicate whether or not such harm is acceptable. Researchers traditionally treat these judgments as a proxy for underlying psychological processes: utilitarian responses presumably reflect cognitive deliberation, whereas deontological responses presumably reflect affective processing. As we will discuss below, this may be an error.

### 3.2.1 A Problem for the Field

The premise that utilitarian judgments arise from cognitive evaluations of outcomes fails to explain the elevated levels of utilitarian responding observed in people who score high on measures of psychopathic personality traits (Bartels & Pizarro, 2011) or patients with damage to the ventromedial cortex (e.g., Ciaramelli et al., 2007; Mendez et al., 2005), considering that such people demonstrate affective deficits, but no cognitive advantages (Koenigs et al., 2007). Therefore, it is possible that these participants’ utilitarian judgments reflect weak aversion to harm rather than strong motivation to maximize outcomes. Traditional methods cannot distinguish between these possibilities.

Traditional methods also fail to capture a central feature of utilitarianism—its emphasis on maximizing good consequences. Philosophical conceptualizations of utilitarianism

entail decision-making focused on maximizing outcomes regardless of whether doing so requires dealing harm (Kagan, 1998; Mill, 1861/1998). Utilitarianism therefore mandates prosocial behavior, such as donating one's disposable income to charity. It is currently unclear whether the psychological processes that motivate utilitarian judgments on moral dilemmas also motivate the prosociality mandated by utilitarian ethics. We employed process dissociation to examine this question as well.

### 3.2.2 Process Dissociation as a Solution

It remains unclear why individuals higher in psychopathic traits prefer overt utilitarian judgments, because more than one process contributes to overt responses to moral dilemmas, an issue called the non-independence error (Conway & Gawronski, 2013). Moral dilemmas, by design, require a tradeoff between utilitarian and deontological options. Every time a person makes a tradeoff of one (moral or amoral) good for another, their responses are multiply-determined. Thus, overt judgments do not cleanly map onto one construct; each judgment results from the strength of both utilitarian and deontological inclinations, as well as other processes. Accordingly, overt 'utilitarian' judgments need not reflect strong utilitarian inclinations; they may instead reflect reduced deontological inclinations, or increased strength of non-moral processes.

In order to overcome the non-independence error, Conway and Gawronski (2013) adapted a technique—process dissociation (PD)—to estimate the independent contributions of two processes when both contribute to overt performance. Although originally developed to disentangle conflated memory processes (Jacoby, 1991), process dissociation is content-agnostic (for a review, see Payne & Bishara, 2009). Process dissociation entails comparing performance on incongruent tasks, where processes are theorized to conflict, with performance on congruent tasks, where processes are theorized to lead to the same response. In the case of moral dilemmas, incongruent dilemmas correspond to traditional, high-conflict moral dilemmas where causing some degree of harm will lead to greater positive outcomes (e.g., Koenigs et al., 2007). Congruent “dilemmas” are those in which harm no longer leads to greater positive outcomes—scenarios in which deontology and utilitarianism ought to lead to the same response (avoid harm).



Consider, for example, the *torture dilemma*, where participants must decide whether it is acceptable to torture a man in order to prevent an explosion. In the incongruent version, the explosion will kill people, so utilitarianism entails accepting torture in order to save lives (maximize outcomes), whereas deontology entails rejecting torture (as harm is always wrong according to deontology). Conversely, in the congruent version of the torture dilemma, the explosion is a messy but harmless paint bomb; now utilitarianism and deontology both entail rejecting torture, as torture no longer maximizes overall welfare.

To the degree that participants experience deontological inclinations, they should reject harm on both incongruent and congruent dilemmas; to the degree that participants experience utilitarian inclinations, they should reject harm on congruent but not incongruent dilemmas. Based on each participants' pattern of responses across both types of dilemmas, process dissociation algebraically derives two parameters that correspond to the degree to which each participant rejected harm on both congruent and incongruent dilemmas (the deontology parameter), and the degree to which participants rejected harm on congruent but not incongruent dilemmas (the utilitarian parameter). Conway and Gawronski (2013) found that the deontology parameter uniquely predicted empathic concern and was enhanced when harm was made salient, whereas the utilitarian parameter uniquely predicted need for cognition and was impaired by a cognitive load manipulation. The parameters themselves were uncorrelated.

### 3.2.3 The Current Work

In Study 1, we employed Conway and Gawronski's process dissociation technique in order to clarify the relations between deontology, utilitarianism, and antisocial personality traits. This analysis allowed us to determine whether the impact of antisocial personality traits on overt utilitarian judgments is driven by increased utilitarian inclinations or reduced deontological inclinations (or both) among people with antisocial personality traits. In Study 2, we examined the association between the utilitarian and deontology process dissociation parameters and prosocial motivation. If measurements of utilitarianism as captured by dilemma studies correspond to philosophical conceptions of

utilitarianism (e.g., Mill, 1861/1998), then the utilitarian parameter ought to positively relate to prosociality.

### 3.3 Study 1

Study 1 examined whether the relation between utilitarian responding and morally questionable personality traits identified by Bartels and Pizarro (2011) is driven by an enhanced endorsement of utilitarianism or a reduced sensitivity to deontological considerations, or both.

#### 3.3.1 Method

**Participants.** Eighty-nine participants (49 male, 40 female) were recruited for payment via www.mturk.com (see Buhrmester, Kwang, & Gosling, 2011): 67 Caucasian, 3 Black, 2 Aboriginal, 10 Asian, and 7 other ( $M_{\text{age}}=34.07$ ,  $SD=13.03$ ).

**Procedure and materials.** Participants read a series of moral dilemmas and completed a set of individual-difference scales (described below).

**Moral dilemmas.** We employed the 10 moral dilemmas used by Conway and Gawronski (2013), each with one *congruent* and one *incongruent* version, in the same fixed random order. Each dilemma described harmful action that would achieve a particular outcome. Participants indicated whether the harmful action was *appropriate* or *inappropriate* (see Greene et al., 2001). Incongruent versions correspond to traditional, high-conflict moral dilemmas (Koenigs et al., 2007), in that action involves causing some degree of harm (e.g., kill one person) to achieve a more positive outcome (e.g., save multiple people). Conversely, congruent versions involve causing some degree of harm (e.g., kill one person) to achieve a less positive outcome (e.g., save no one).

**Individual difference scales.** Following the dilemma task, participants responded to the 72 individual-difference items employed by Bartels and Pizarro (2011) on 7-point scales ranging from *strongly disagree* to *strongly agree*. The order of items was randomized for each participant. The items comprised a 30-item psychopathy scale (SRP-III; Paulhus, Neumann, & Hare, in press), e.g., *I never feel guilty over hurting others* ( $\alpha=.92$ ), a 20-

item Machiavellianism scale (Mach-IV; Christie & Geis, 1970), e.g., *Never tell anyone the real reason you did something unless it is useful to do so* ( $\alpha=.83$ ), an 18-item No Meaning scale (Kunzendorf, Moran, & Gray, 1995), e.g., *All strivings in life are futile and absurd* ( $\alpha=.95$ ) and a 10-item Social Desirability scale (MC-1; Strahan & Gerbasi, 1972), e.g., *I'm always willing to admit it when I make a mistake* ( $\alpha=.67$ ). Finally, participants reported their age, gender, and ethnicity.

### 3.3.2 Results and Discussion

We calculated overt ‘utilitarian’ dilemma judgments by summing the number of times participants indicated that harmful action was acceptable across the ten incongruent (high-conflict) moral dilemmas (following Greene et al., 2001). Note that overt judgments may be described as either ‘utilitarian’ or ‘deontological’ depending on whether one tallies acceptance or rejection of harm. Examining overt judgments, we replicated the findings reported by Bartels and Pizarro (2011): people who made more ‘utilitarian’ judgments also scored higher on psychopathy, Machiavellianism, and meaninglessness (see Table 3). As discussed above, these findings remain ambiguous because overt judgments reflect both utilitarianism and deontology. Therefore, these findings might indicate positive relations between utilitarianism and antisocial personality traits, or they might indicate negative relations between deontology and antisocial personality traits. Process dissociation can disambiguate these possibilities.

We calculated utilitarian and deontology process dissociation parameters according to the metric provided by Conway and Gawronski (2013), and examined the correlations between the utilitarian and deontological PD parameters and each personality variable. Consistent with Conway and Gawronski, overt utilitarian judgments correlated positively with the utilitarian PD parameter and negatively with the deontological PD parameter—these correlations were moderate to strong (Cohen, 1988)—suggesting that the PD parameters and overt dilemma judgments are tapping related, but not redundant, constructs (see Table 3). Moreover, the PD parameters were uncorrelated with one another, replicating Conway and Gawronski, and further corroborating the independence of these processes (Greene, 2007).

	<b>Overt 'Utilitarian' Judgments</b>	<b>Utilitarian PD Parameter</b>	<b>Deontology PD Parameter</b>
PD Utilitarianism	.44***		
PD Deontology	-.78***	.15	
Machiavellianism	.41***	.02	-.49***
Psychopathy	.41***	-.10	-.54***
No Meaning	.30**	-.16	-.43***
Social Desirability	-.21*	-.09	.21*
Gender (1 = male, 2 = female)	-.28**	-.01	.30**

*Note:* \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Table 3: Correlations between Overt 'Utilitarian' Judgments, Process Dissociation Utilitarianism Scores, Process Dissociation Deontology Scores, and Personality Measures in Study 1 (N = 89).

When measured via process dissociation, the association between utilitarianism and antisocial personality traits disappeared: none of the correlations with the utilitarian PD parameter reached significance. Conversely, all three personality measures correlated negatively with the deontology PD parameter.<sup>1</sup> These findings clarify the ambiguity in previous findings regarding increased overt ‘utilitarian’ judgments among people high in antisocial personality traits (Bartels & Pizarro, 2011) or patients with damage to the VMPFC (e.g., Koenigs et al., 2007). The current data suggest that people scoring higher in psychopathy, Machiavellianism, and meaninglessness were not experiencing *stronger* utilitarian moral inclinations. Rather, they were experiencing *weaker* deontological moral inclinations. In other words, these participants higher in antisocial personality traits cared less about the prohibition against causing the death of innocents. Accordingly, they *appeared* to make more utilitarian judgments, but this finding is better described as making fewer deontological judgments. Therefore, overt ‘utilitarian’ responses on moral dilemmas may reflect genuine utilitarian inclinations, but they may also reflect other processes (e.g., selfishness), especially when competing deontological inclinations are relatively weak, as in the case of psychopaths or people with damage to the VMPFC (e.g., Koenigs et al., 2007; Saver & Damasio, 1991).

### 3.4 Study 2

Although Study 1 demonstrated that the association between utilitarian dilemma judgments and antisocial personality traits was driven by reduced deontological rather than enhanced utilitarian inclinations, the question remains whether utilitarian inclinations map onto traditional philosophical conceptions of utilitarianism (e.g., Kagan, 1998; Mill, 1861/1998), which emphasize the maximization of outcomes despite the causing of harm. If so, utilitarian responses ought to correlate positively with various measures of prosocial motivation. Study 2 examined whether this is the case.

#### 3.4.1 Method

**Participants.** Ninety-one participants (36 male, 54 female, 1 unreported) were recruited for payment via [www.mturk.com](http://www.mturk.com): 69 Caucasian, 10 Black, 4 Asian, 3 Latino, 2 other ethnicity, and 3 unreported ( $M_{\text{age}} = 36.74$ ,  $SD = 12.30$ ).

**Procedure and materials.** Participants completed several measures of prosocial motivation and then responded to the same congruent and incongruent moral dilemmas in the same fixed random order as Study 1.

**Helping Scenarios.** Participants read four brief vignettes where the protagonist requested mundane assistance, such as delivering a parcel 15 minutes out of your way (Conway & Peetz, 2012). Participants indicated the likelihood that they would help each protagonist on 7-point scales from *not at all* to *very much* ( $\alpha = .64$ ).

**Willingness to Volunteer Scale.** Participants reported their willingness to volunteer for a worthy cause (Devoe & Pfeffer, 2007) by rating their agreement with five items such as *Volunteering is a worthwhile use of my time even if I do not get paid* on 7-point scales ranging from *strongly disagree* to *strongly agree* ( $\alpha = .90$ ).

**Charity Donation Task.** Participants were given an option to enter a dovert for US\$50, and they indicated, if they won, how much money they would like donated on their behalf to four charities (UNICEF, The American Red Cross, The World Wildlife Federation, and a charity of their own choosing, Conway & Peetz, 2012). Sixty-six people entered the draw; 25 refused.

**Morality of Prosocial Behavior.** Participants responded to nine prosocial dilemmas (eight adapted from Unger, 1996; one original) where protagonists considered actions that benefit others despite a cost to oneself (see Appendix 3A). Participants indicated the morality of each action on 7-point semantic differential scales ranging from *morally bad* to *morally good*. For example, in the *yacht* dilemma, participants indicated the morality of lying to a guard, then stealing and damaging their boss's yacht in order to save a drowning woman. We reverse-coded responses to the *tsunami*, *envelope*, *resume*, *lost wallet*, and *stock tip* dilemmas so that higher scores indicated approval of utilitarian action.

We used three strategies to measure approval of prosociality. First, we summed approval of prosocial actions across all dilemmas ( $\alpha = .51$ ). Second, we conducted an exploratory factor analysis, which suggested that prosocial dilemmas loaded onto three separate

factors with eigenvalues greater than 1 (see Table 4 for each dilemma's factor loadings on each varimax rotated factor). The factors were A) altruism dilemmas (yacht, lost little girl, architect, resume,  $\alpha = .68$ ) where the actor must choose between prosocial versus selfish options (e.g., risking one's life to save one's boss), B) honesty dilemmas (account, wallet, stock tip reverse-coded,  $\alpha = .61$ ) where the actor must decide whether to be honest or dishonest, and C) charity dilemmas, where the actor must decide whether or not to donate to a charity that will help many distant victims (envelope & tsunami,  $\alpha = .67$ ). Third, we examined approval for each dilemma separately.

### 3.4.2 Results and Discussion

As in Study 1, we calculated overt 'utilitarian' judgments and both process dissociation parameters, and examined the correlations between these variables and each measure of prosocial motivation (see Table 5). Once again, overt 'utilitarian' judgments correlated positively with the utilitarian PD parameter and negatively with the deontological PD parameter, but the parameters themselves were uncorrelated.

As in Study 1, analyzing overt dilemma responses makes utilitarianism appear morally questionable, as 'utilitarian' judgments were associated with reduced volunteerism and increased acceptance of dishonesty. However, these findings may reflect reduced prosociality among people who made more utilitarian judgments, or increased prosociality among people who made more deontological judgments. Process dissociation can disambiguate these possibilities.

When assessed via process dissociation, the negative relation between utilitarianism and volunteerism disappeared, as did the relation between utilitarianism and acceptance of dishonesty.<sup>2</sup> Instead, there was a significant positive relation between deontology and volunteerism, and deontology and honesty. These findings clarify that utilitarian inclinations do not reduce motivation to volunteer or be honest—rather, deontological inclinations increase motivation to volunteer and be honest. Moreover, process dissociation revealed an association between deontology and prosocial motivation on mundane helping scenarios (specifically, the account, wallet, and stock account

<b>Prosocial Dilemma</b>	<b>Altruism Factor</b>	<b>Honesty Factor</b>	<b>Charity Factor</b>
Yacht Dilemma <sup>a</sup>	<b>.75</b>	.16	-.12
Lost Little Girl Dilemma <sup>a</sup>	<b>.82</b>	-.06	.17
Architect Dilemma <sup>a</sup>	<b>.61</b>	-.22	.20
Resume Dilemma <sup>a</sup>	<b>.57</b>	-.35	.21
Account Dilemma <sup>b</sup>	.15	<b>.88</b>	.01
Wallet Dilemma <sup>b</sup>	-.41	<b>.62</b>	.22
Stock Tip Dilemma <sup>b</sup>	.47	<b>-.58</b>	.36
Envelope Dilemma <sup>c</sup>	.06	.01	<b>.83</b>
Tsunami Dilemma <sup>c</sup>	.14	.01	<b>.84</b>

Table 4: Factor Loadings of Judgments in Prosocial Dilemmas on Three Rotated Factors with Eigenvalues Greater than One in Exploratory Factor Analysis, Study 2(N = 91).

*Note:* We selected an arbitrary criterion that judgments loading  $> \pm .5$  on a given factor meant that item was included in that factor. No item loaded on more than one factor. Dilemmas that share the same subscript comprise the factor with that subscript. The Stock Tip Dilemma loaded negatively on honesty factor, so it was recoded such that higher scores on all three honesty dilemmas reflected moral approval of dishonesty. However, there is an important difference: in the account and wallet dilemmas, honest answers benefit authority figures at a cost to victims, whereas in the stock tip dilemmas honesty harms an authority figure and benefits society—so dishonesty is not strictly utilitarian in this combination.



	<b>‘Utilitarian’ Judgments</b>	<b>Utilitarian PD Parameter</b>	<b>Deontology PD Parameter</b>
PD Deontology	-.75***		
PD Utilitarianism	.64***	-.01	
Scenario Help	-.14	.13	.28**
Volunteerism	-.21*	.12	.38***
Charity Donations	.14	.00	-.16
Gender (1 = male , 2 = female)	-.41***	-.22*	.38***
Overall Dilemma Help	.10	.24*	.03
Altruism Dilemmas <sup>a</sup>	.01	.28**	.17
Honesty Dilemmas <sup>b</sup>	.28*	-.05	-.48***
Charity Dilemmas <sup>c</sup>	-.03	.13	.09
Yacht Dilemma <sup>a</sup>	.12	.24*	-.03
Lost Little Girl Dilemma <sup>a</sup>	-.003	.27**	.18
Architect Dilemma <sup>a</sup>	-.13	.10	.23*
Resume Dilemma <sup>a</sup>	.03	.21	.14
Account Dilemma <sup>b</sup>	.31**	.04	-.37**
Wallet Dilemma <sup>b</sup>	.14	-.11	-.29**
Stock Tip Dilemma <sup>b</sup>	.16	-.06	-.29**
Envelope Dilemma <sup>c</sup>	.01	.09	.02
Tsunami Dilemma <sup>c</sup>	-.05	.12	.14

Table 5: Correlations between Traditional Bipolar Utilitarianism Scores, Process Dissociation Utilitarianism Scores, Process Dissociation Deontology Scores, and Prosocial Dependent Measures in Study 2 (N = 91).

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

*Note:* Dilemmas that share the same subscript comprise the factor with that subscript

*Note:* Stock Tip Dilemma loaded negatively on honesty factor, so it was recoded such that higher scores on all three honesty dilemmas reflected moral approval of dishonesty.

However, there is an important difference: in the account and wallet dilemmas, honest answers benefit authority figures at a cost to victims, whereas in the stock tip dilemmas honesty harms an authority figure and benefits society—so dishonesty is not strictly utilitarian in this combination.

dilemmas) that was obscured in overt dilemma judgments. These findings suggest that deontological inclinations motivate prosociality in mundane situations, such as volunteering or doing modest favors for others.

Process dissociation also revealed associations between utilitarianism and prosocial motivation that were obscured using overt judgments. Specifically, there was a positive relation between the utilitarian parameter and approval of costly prosocial actions overall. This finding was driven by the subset of dilemmas that loaded on the altruism factor (particularly, the yacht and lost little girl dilemmas). These findings suggest that utilitarian inclinations motivate prosociality in more extreme situations where someone's life may be at risk. Hence, utilitarianism and deontology each appear to motivate prosociality under different circumstances.

Surprisingly, no relations emerged between either parameter and charity donations, or responses on charity dilemmas. Excluding participants who refused to enter the doverting did not alter this finding: none of the correlations reached significance (all  $r$ s < .10, all  $p$ s > .10). Perhaps utilitarianism would have predicted donations if we clarified that donations would save lives, or deontology would have predicted if we enhanced the emotional salience of victims. Future work could determine whether this is the case.

### 3.5 General Discussion

These two studies provide a clear account regarding the utilitarian and deontological inclinations that underpin judgments on traditional moral dilemmas. On the one hand, overt 'utilitarian' judgments in Study 1 correlated with psychopathy, Machiavellianism, and meaninglessness (replicating Bartels, & Pizarro, 2011), as well as dishonesty and reduced volunteerism in Study 2. However, a process dissociation analysis revealed that the increased overt 'utilitarian' responses expressed by people scoring high in antisocial personality traits in Study 1 were actually a function of reduced deontological inclinations, rather than increased utilitarian inclinations. Moreover, although overt 'utilitarian' judgments in Study 2 were associated with reduced volunteerism, process dissociation clarified that this finding was not because utilitarian inclinations are associated with reduced volunteerism; rather, deontological inclinations were associated

with increased volunteerism. In addition, utilitarian inclinations positively correlated with motivation to help on prosocial dilemmas—particularly those that involve making a substantial personal sacrifice to help another—whereas deontological inclinations were related to willingness to volunteer, mundane prosociality, and a desire for honesty (even honesty at a personal cost). Hence, utilitarianism and deontology each predict different kinds of prosociality.

In Study 2, utilitarian inclinations were associated with stronger prosocial motivation in extreme, life-or-death scenarios where victims faced extreme harm (e.g., the yacht dilemma), whereas deontological inclinations were associated with prosocial motivation in mundane situations (e.g., delivering a parcel, volunteering). This suggests that mundane prosocial motivations might be stimulated by deontological inclinations—theorized to involve empathy for/perspective-taking of a person in need—whereas extreme prosocial motivations are stimulated by utilitarian inclinations, theorized to involve a cognitive strategy of maximizing overall welfare. Future work should consider this possibility more systematically.

The current findings suggest an explanation for why people high in antisocial personality traits (Bartels & Pizarro, 2011) or damage to the VMPFC (Koenigs et al., 2007) make more utilitarian judgments. When faced with dilemmas that pit emotional reactions against deliberative reasoning, such people experience cognitive processing similar to other participants, coupled with reduced emotional processing. As a result, such decisions are improperly characterized as ‘more utilitarian’ than controls, when they are more accurately described as ‘less deontological.’ Accordingly, researchers should use caution when interpreting overt dilemma judgments: just because a given participant made many utilitarian judgments does not guarantee that they did so *because* they have strong utilitarian inclinations. Weak deontological inclinations can also produce a pattern of responding on moral dilemmas that erroneously appear to reflect strong utilitarian concerns.

Note that these findings provide enhanced support for the dual process model than was previously available (Greene, 2007): if deontological judgments are driven by affective

reactions to harm, it makes sense that people who experience reduced affectivity are less motivated to make deontological judgments (rather than more motivated to make utilitarian judgments). Finally, these findings highlight the utility of process dissociation for clarifying the nature of moral inclinations underpinning overt dilemma judgments, which conflate the impact of each inclination.

In sum, traditional measurement techniques have led researchers to conflate overt judgments on moral dilemmas with the inclinations underlying those judgments. The result of this error is theoretical confusion, such as unwittingly labeling people scoring high in antisocial personality traits as ‘more utilitarian.’ In the current work, we employed process dissociation to clarify that (a) the link between overt ‘utilitarian’ moral dilemma judgments and antisocial personality traits is spurious (it is not driven by increase utilitarian inclinations, but rather, by reduced deontological inclinations), and (b) that utilitarianism is positively related to some kinds of prosocial motivation (extreme circumstances), whereas deontology is related to other kinds of prosocial motivation (mundane circumstances). Researchers should exercise caution when interpreting overt dilemma judgments; a higher proportion of one kind of judgment need not imply the presence of a corresponding moral inclination.

## References

- Amazon (2011). <https://requester.mturk.com/> Retrieved on July 3, 2012.
- Baron, J. (2011). Utilitarian emotions: Suggestions from introspection. *Emotion Review*, 3, 286-287. doi:10.1177/1754073911402377
- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, 121, 154-161. doi:10.1016/j.cognition.2011.05.010
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5. doi:10.1177/1745691610393980
- Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. New York: Academic Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second edition). New Jersey: Lawrence Erlbaum Associates.
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, 104, 216-235. doi:10.1037/a0031021
- Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, 38, 907-919. doi:10.1177/0146167212442394
- Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2, 84-92. doi:10.1093/scan/nsm001
- Devoe, S. E., & Pfeffer, J. (2007). Hourly payment and volunteering: The effect of organizational practices on decisions about time use. *Academy of Management Journal*, 50, 783-798.
- Ginges, J., Atran, S., Medin, S., & Shikaki, K. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences*, 104, 7357-7360. doi:10.1073/pnas.0701768104
- Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, 11, 322-323. doi:10.1016/j.tics.2007.06.004

- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105-2108.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513-541.
- Kagan, S. (1998). *Normative Ethics*. Westview Press: Boulder, CO.
- Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, 7, 393-402.
- Kant, I. (1785/1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446, 908-911. doi:10.1038/nature05631
- Kunzendorf, R. G., Moran, C., & Gray, R. (1995). Personality traits and reality-testing abilities, controlling for vividness of imagery. *Imagination, Cognition, and Personality*, 15, 113-131.
- Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology*, 18, 193-197.
- Paulhus, D.L., Neumann, C.F., & Hare, R.D. (in press). *Manual for the self-report psychopathy scale*. Toronto: Multi-Health Systems.
- Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, 20, 272-314. doi:10.1080/10463280903162177
- Saver, J. L. & Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, 29, 1241-1249. doi:10.1016/0028-3932(91)90037-9
- Strahan, R. & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology*, 28, 191-193.
- Unger, P. (1996). *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press.

## Notes

<sup>1</sup> We replicated the finding that men scored higher on psychopathy than women (although no gender differences emerged for Machiavellianism or meaninglessness), and social desirability correlated negatively with all three personality measures. Moreover, men made more overt utilitarian judgments than women, an effect that PD revealed to be driven by increased deontological (but not utilitarian) concern among women. Additionally, people scoring higher in social desirability made fewer overt ‘utilitarian’ judgments than people less concerned with social desirability, an effect that PD revealed to be driven by increased deontological (but not utilitarian) concern among those scoring higher in social desirability.

To examine whether the relations between antisocial-personality traits and moral inclinations held above and beyond these factors, we separately regressed a) overt ‘utilitarian’ dilemma judgments, b) the utilitarian PD parameter, and c) the deontology PD parameter on each of our (centered) predictor variables while controlling for social desirability and gender (see Table 6). All three personality variables remained significant positive predictors of overt utilitarian judgments in this analysis. Moreover, all three personality variables remained significant negative predictors of the deontology PD parameter while controlling for social desirability and gender, and they did not become positive predictors of the utilitarian PD parameter in this analysis. Therefore, the relations between antisocial personality traits and reduced deontological inclinations cannot be attributed to gender or social desirability concerns.



<b>Personality Measure</b>	<b><math>\beta</math> Overt 'Utilitarian'</b>	<b><math>\beta</math> Social Desirability</b>	<b><math>\beta</math> Gender</b>
Machiavellianism	.39**	.02	-.22*
Psychopathy	.33*	-.04	-.13
No Meaning	.23*	-.14	-.22*
<b>Personality Measure</b>	<b><math>\beta</math> Utilitarian PD Parameter</b>	<b><math>\beta</math> Social Desirability</b>	<b><math>\beta</math> Gender</b>
Machiavellianism	-.05	-.88	-.01
Psychopathy	-.24	-.20	-.09
No Meaning	-.20	-.13	-.03
<b>Personality Measure</b>	<b><math>\beta</math> Deontology PD Parameter</b>	<b><math>\beta</math> Social Desirability</b>	<b><math>\beta</math> Gender</b>
Machiavellianism	-.49***	-.08	.23*
Psychopathy	-.54***	-.06	.07
No Meaning	-.36***	.11	.22*

Table 1: Regressing Utilitarian Decisions, PD Utilitarianism Scores, and PD Deontology Scores on Machiavellianism, Psychopathy, and No Meaning, Controlling for Social Desirability and Gender in Study 1 (N = 89).

*Note:* \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . Gender coded 1 = male, 2 = female

<sup>2</sup>As in Study 1, men were more likely to make utilitarian judgments than women, but this time a process dissociation analysis indicated that this relation obtained both due to increased deontological inclinations among women and increased utilitarian inclinations among men. In order to determine whether our findings obtained above and beyond gender, we separately regressed a) overt 'utilitarian' dilemma judgments, b) the utilitarian PD parameter, and c) the deontology PD parameter on each of our (centered) predictor variables while controlling for gender (see Table 7).

Overt 'utilitarian' judgments were significantly predicted by gender,  $\beta = -.37$ ,  $p < .001$ , and marginally predicted by charity donations,  $\beta = .18$ ,  $p = .078$ , and approval of action in altruism dilemmas,  $\beta = .20$ ,  $p = .069$ . Separately, we conducted identical analyses on the utilitarian and deontological PD parameters. Approval of action in altruism dilemmas remained a significant predictor of the utilitarian parameter,  $\beta = .30$ ,  $p = .011$ , as did gender,  $\beta = -.27$ ,  $p = .014$ . No other effects were significant (all  $ps < .1$ ). Volunteering significantly predicted the deontology parameter,  $\beta = .23$ ,  $p = .036$ , as did disapproval of dishonest action in the honesty dilemmas,  $\beta = -.25$ ,  $p = .018$ , and gender,  $\beta = -.37$ ,  $p < .001$ . No other effects were significant (all  $ps < .1$ ).

The results of this analysis were similar to the correlational analysis, with the exception that the relation between utilitarian responses and volunteerism no longer reached significance. Otherwise, the relations between overt dilemma decisions, process dissociation parameters, and prosocial variables remained significant. Therefore, these results cannot be attributed to the influence of gender.

Finally, we conducted a multiple regression analysis where we simultaneously regressed utilitarian decisions on all prosocial variables and gender. In this analysis, only gender significantly predicted overt 'utilitarian' dilemma judgments,  $\beta = -.37$ ,  $p < .001$ , but charity donations,  $\beta = .18$ ,  $p = .078$ , and altruism dilemmas,  $\beta = .20$ ,  $p = .069$ , marginally predicted utilitarian judgments. No other effects were significant (all  $ps < .1$ ).

<b>Prosocial Measure</b>	<b><math>\beta</math> Overt 'Utilitarian' Dilemma Judgments</b>	<b><math>\beta</math> Gender</b>
Scenario Help	-.14	-.41***
Volunteerism	-.17	-.38***
Charity Donations	.14	-.41***
Overall Dilemma	.10	-.41***
Altruism Dilemmas	.04	-.41***
Honesty Dilemmas	.21*	-.35**
Charity Dilemmas	-.02	-.41***
<b>Prosocial Measure</b>	<b><math>\beta</math> Utilitarian PD Parameter</b>	<b><math>\beta</math> Gender</b>
Scenario Help	.13	-.22*
Volunteerism	.15	-.25*
Charity Donations	.01	-.22*
Overall Dilemma	.25*	-.24*
Altruism Dilemmas	.31**	-.26*
Honesty Dilemmas	-.11	-.25*
Charity Dilemmas	.13	-.23*
<b>Prosocial Measure</b>	<b><math>\beta</math> Deontology PD Parameter</b>	<b><math>\beta</math> Gender</b>
Scenario Help	.28**	.37***
Volunteerism	.34***	.32**
Charity Donations	-.14	.38***
Overall Dilemma	.03	.38***
Altruism Dilemmas	.16	.26**
Honesty Dilemmas	-.39***	.28**
Charity Dilemmas	.09	.37***

Table 2: Regressing Utilitarian Decisions, PD Utilitarianism Scores, and PD Deontology Scores on Scenario Help, Volunteerism, Donations, Dilemma Help, help on Altruism, Honesty, and Charity Dilemmas, and Gender in Study 2 (N = 91).

## Appendix 3A: Prosocial Dilemmas

Please read each of the following scenarios and tell us how morally good or bad the action is, in your opinion.

### **Account**

You work for a large accounting firm whose clients include a certain billionaire. You know he gives a lot of money to several fashionable charities, but does hardly anything to aid the world's neediest people. You find out that you can decrease by only a million dollars, the billionaire's huge account without it ever being noticed: Through a sequence of many small anonymous contributions, the million will go to an international aid organization and, as a result, ten thousand fewer children will die in the next few months. You shift the funds and ten thousand more children are saved.

### **Tsunami**

You're visiting a friend's office who does research on beach erosion in a region of Asia. He has a webcam overlooking the beach, so he can monitor the site while in the U.S. He is showing web feed, when you suddenly see a wave approaching the beach, resulting from a tsunami. You watch, as the waves smash the village of a few thousand people, destroying property and killing many people instantly. Later that day, you get an email from an international aid organization that you trust. It says a donation of \$100 made through its website will be deployed almost instantaneously to provide lifesaving food and medical care to ten people. However, you decide not to donate money.

### **Yacht**

You have a summer job working on the waterfront estate of a billionaire. Through binoculars, you see a woman out in the waves who will drown unless you help her now. To rescue her, you must steal your boss' yacht, which is worth millions of dollars. To steal the yacht, you will have to lie to the guard watching the dock. And if you go, you'll have to pass through a channel where the yacht will be damaged. You're not rich, so you won't be

able to help much with the repair bill that, even after insurance, will be over a million dollars. Still, you take the yacht and save the woman.

### **Envelope**

You receive a letter from an international aid organization. You respect the organization for the work they do and for their reputation of honesty, efficiency in delivering aid and commitment. The letter tells you about an epidemic in Africa that is causing many children to die from dehydration. A simple treatment called “oral rehydration therapy” can save these children’s lives. You read that a \$100 contribution will save the lives of thirty children. An addressed envelope is enclosed, making it easy to send your check. But you throw the envelope in the trash, and instead of living many years, thirty more children soon die than would have if you had contributed \$100.

### **Lost Little Girl**

You are trespassing on a farmer’s private property in order to harvest some rare plants that grow there. It is well known that the farmer does not like people harvesting plants on his property. He has put up large signs warning that trespassers will be prosecuted. In a dense forested area you find the farmer’s young daughter alone, lost, and crying. You are so far from civilization that she is unlikely to be found by any search parties for days. You decide to lead the little girl back to her father even though you will have to admit that you were trespassing on his land. Is it appropriate for you to save the little girl even if it means you might be prosecuted for trespassing?

### **Architect**

You are a young architect visiting one of your construction sites with your boss. Your boss is a well-liked individual who has many friends in the business and community. One day at a building you are inspecting he slips and starts to fall off the roof. You could reach out and pull him back to safety, but there is a chance you might start to fall too. Is it appropriate for you to lean out to save your boss if it means risking your life in the process?

**Resume**

You have a friend who is looking for a job. He looks at his resume and thinks that it is not quite strong enough to get the position that he really wants. He considers putting some false information on his resume in order to make it more impressive. By doing this he may manage to get hired, beating out several candidates who are actually more qualified than he. In the end, though, he decides to leave it the way it is even if that means he is less likely to get hired. Is it appropriate for your friend to be truthful on his resume even if he is less likely to get the position he wants?

**Lost Wallet**

You are walking down the street when you come across a wallet lying on the ground. You open the wallet and find that it contains over a thousand dollars in cash as well the owner's driver's license. You recognize the owner of the wallet as a businessman who is famously rich and greedy. He recently fired many people in town from their jobs. You consider sending the wallet back to the owner without the cash and distributing it among those people recently fired by the business man. But the owner of lost property deserves to get that property back intact, so you consider contacting the businessman to give him back everything. Is it appropriate for you give back the money in the wallet if it could have helped people fired by the wallet's owner?

**Stock Tip**

You are a management consultant working on a case for a large corporate client. Your boss has access to confidential information that would be very useful to investors. One night you catch him leaking information in return for large sums of money. Releasing information in this way is strictly forbidden by federal law. Your first instinct is to tell everything to federal investigators. But then you realize that if you did so, you would probably lose your job and never be hired again in the industry. Also your boss would be very angry with you when he got out of jail. Is it appropriate for you to inform federal investigators if it means you will likely lose your job?

## Chapter 4

### 4 Conclusions General and Discussion

Together, the three studies reported in Chapter 2 and two studies reported in Chapter 3 provide improved understanding of the psychology of moral judgments. Chapter 2 presented three studies that validated the application of Jacoby's (1991) process dissociation procedure to moral judgments, thereby allowing for independent estimates of deontology and utilitarianism and overcoming the non-independence error that plagues other moral dilemma research. By measuring each inclination separately, it becomes possible to determine the degree to which each contributes to moral judgments. The utility of this technique was further documented in Chapter 3, where process dissociation clarified theoretical confusions in the field, specifically the relations between deontology, utilitarianism, antisocial personality traits, and prosociality.

#### 4.1 Results

Let's begin with a review of the results of each study in turn, beginning with Conway and Gawronski (2013) Studies 1, 2, and 3 in Chapter 3 and continuing through Conway, Bartels, and Pizarro (under review) Studies 1 and 2 in Chapter 3.

##### 4.1.1 Results: Chapter 2, Study 1

In Chapter 2, three studies examined the validity of process dissociation for measuring the processes involved in moral dilemma judgments. The predominant dual-process model of moral judgments (Greene, 2007) postulates that *deontological* moral judgments—refusing to cause harm even when harm will lead to the best overall outcome—are driven by affective reactions to harm, whereas *utilitarian* moral judgments—acting to produce the best overall outcome, even if that entails causing harm—are driven by cognitive evaluations of outcomes. Although a large and growing body of research supports this theory (e.g., Bartels & Pizarro, 2011; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Koenigs et al., 2007; Suter & Hertwig, 2010), researchers have typically measured deontology and utilitarianism by pitting one against the other. Therefore, it remains unclear whether overt moral judgments are

actually driven by one process, the other process, or a combination of the two; a single-process account cannot be ruled out.

Process dissociation (Jacoby, 1991) allows for the independent assessment of each process by pitting deontology against utilitarianism in some dilemmas, whereas in other dilemmas deontology accords with utilitarianism. By assessing responses on both types of dilemmas, it is possible to mathematically isolate and estimate each process: the deontology and utilitarian parameters. These parameters may be compared with overt judgments, one another, and third variables, in order to clarify the underlying relations between these constructs. In addition, it is also possible to manipulate variables that accord with theoretical conceptions of each process, in order to determine whether these manipulations affect the parameters independently of one another. If the parameters correlate sensibly with overt judgments and third variables, and are selectively impacted by theoretically-derived manipulations, this would suggest that the parameters are accurately assessing deontology and utilitarianism. The three studies reported in Chapter 2 did exactly this.

Study 1 in Chapter 2 examined the correlations between overt moral judgments, the deontology and utilitarian process dissociation parameters, and a number of individual difference measures: empathy and perspective-taking, need for cognition, faith in intuition, moral identity (internalization), and religiosity. Results suggested that the deontological and utilitarian parameters did, indeed, relate sensibly to overt moral judgments: the more overt “deontological” moral judgments (i.e., fewer “utilitarian” overt moral judgments) participants made, the higher they scored on the deontological parameter, whereas the fewer overt “deontological” moral judgments (i.e., more “utilitarian” overt moral judgments) participants made, the higher they scored on the utilitarian parameter. Moreover, the two parameters themselves were not significantly correlated, suggesting that deontology and utilitarianism do, indeed, reflect independent constructs, in line with the dual-process model (Greene, 2007).

In terms of relations with third variables, “deontological” overt moral judgments (i.e., reduced “utilitarian” overt moral judgments) were positively correlated with empathic



concern and perspective-taking. The deontological parameter also correlated positively with both these constructs, whereas the utilitarian parameter did not significantly correlate with either. This finding is consistent with the dual-process model whereby deontology is primarily driven by an affective reaction to harm—which ought to be stronger in individuals higher in empathy and perspective-taking—whereas utilitarianism is primarily driven by cognitive processes unrelated to these constructs. In addition, “deontological” overt moral judgments (i.e., reduced “utilitarian” overt moral judgments) were negatively correlated with need for cognition. This time, the deontological parameter was unrelated to need for cognition, whereas the utilitarian parameter was positively related. This pattern is consistent with the theoretical position that utilitarianism is driven by cognitive evaluations of outcomes. “Deontological” overt moral judgments (i.e., reduced “utilitarian” overt moral judgments) were also positively correlated with religiosity, and process dissociation again clarified the nature of these relations: the deontological parameter was positively correlated, but the utilitarian parameter was uncorrelated, with this construct. This was an exploratory analysis, so it does not necessarily corroborate or refute the dual-process model per se, but it reinforces the conclusion that the parameters are tapping independent constructs that are conflated in overt moral judgments.

Finally, the most interesting pattern emerged for moral identity: there was no significant correlation between moral identity and overt dilemma judgments; however, both parameters positively correlated with moral identity. This finding has two implications: first, it suggests that because overt moral judgments conflate deontology and utilitarianism, process dissociation can uncover parallel relations with third variables that are cancelled out when these processes are conflated. Second, this finding suggests that *both* deontology and utilitarianism are genuinely moral processes. Without this evidence, it might be possible to assume that utilitarianism is a process entailing morally-neutral calculations; this evidence suggests that utilitarianism is calculation *for moral ends*. To the author’s knowledge, this is the first evidence to suggest there is something genuinely moral about utilitarianism, a proposition which had been in doubt (see Bartels & Pizarro, 2011). Moreover, this conclusion was reinforced by the findings regarding the relation

between utilitarianism and prosocial motivations from Study 2, Chapter 3, which will be discussed below.

In sum, Study 1 in Chapter 2 suggested that the deontological and utilitarian parameters are tracking primarily affective and primarily cognitive processes, respectively, and clarify the relations between overt moral judgments and individual differences by delineating when these relations may be attributed to one process, when to another, and when to both. That said, these findings are merely correlational in nature. Although a regression analysis confirmed that the obtained relations remained when controlling for the relation with both parameters simultaneously, stronger evidence that deontology and utilitarianism are, indeed, distinct processes would be obtained if we manipulated theoretically-relevant variables and observed impacts on one but not the other parameter. Hence, Studies 2 and 3 in Chapter 2 did exactly that.

#### 4.1.2 Results: Chapter 2, Studies 2 & 3

In Study 2, Chapter 2, we manipulated cognitive load. Previous work has suggested that inducing cognitive load (Greene et al., 2008) or time pressure (Suter and Hertwig, 2011) impairs utilitarian responding, which is consistent with the dual-process model, where utilitarian judgments are theorized to result primarily from a cognitive evaluation of outcomes. However, previous work suffers from the non-independence error, so it cannot rule out whether load actually facilitates deontology, or perhaps impacts both deontology and utilitarianism. We found that cognitive load reduced the number of overt utilitarian judgments (i.e., increased the number of overt deontological judgments), but process dissociation clarified the nature of this effect: cognitive load selectively reduced the utilitarian parameter without impacting the deontological parameter. This finding provides the best evidence to date that utilitarianism is, indeed, linked to cognitive evaluations of outcomes, and is independent from deontology.

In Study 3, Chapter 2, we manipulated the salience of harm by including a photo of the victim. Previous work (e.g., Bartels, 2008) has suggested a link between making harm vivid and increased deontological judgments; however, this work suffers from the non-independence error, and therefore cannot rule out that making harm salient may also

reduce utilitarianism or affect both processes. We found that making harm salient did, indeed, increase overt “deontological” judgments (i.e., reduced overt “utilitarian” judgments), but process dissociation clarified the nature of this effect: making harm salient selectively increased the deontological parameter without impacting the utilitarian parameter. These findings provide the best evidence to date that deontology is driven primarily by an affective reaction to harm, and operates independently of utilitarianism.

Together the three studies reported in Chapter 2 validate process dissociation as an improved methodology for independently quantifying deontology and utilitarianism, providing the best evidence to date that they are independent processes, that utilitarianism is linked to cognition and deontology to affect, and that both are truly moral in nature.

#### 4.1.3 Results: Chapter 3, Studies 1 & 2

Having validated process dissociation as an improved tool for measuring deontology and utilitarianism, we now looked to see whether process dissociation could help clarify theoretical ambiguity in the field in Chapter 3. The first problem we examined was that of antisocial personality traits. Several lines of work have linked antisocial personality traits to increased utilitarian responding, whether those traits are measured as individual differences in psychopathy, Machiavellianism, and (life) meaninglessness in the general population (e.g., Bartels & Pizarro, 2011), or whether those traits arise through damage to the ventromedial prefrontal cortex, resulting in a pattern of behavior characterized as *acquired sociopathy* that entails a callous disregard for the wellbeing of others (Koenigs et al., 2007; Saver & Damasio, 1991). Yet, those studies suffer from the non-independence error, and therefore cannot determine whether people with antisocial personality traits experience stronger utilitarian inclinations, weaker deontological inclinations, or some combination thereof. In Study 1, Chapter 3, we attempted to replicate Bartels and Pizarro while employing process dissociation in order to clarify the nature of the relations between antisocial personality traits and utilitarianism.

The results of Study 1, Chapter 3 replicated those of Bartels and Pizarro when examining overt judgments: people scoring higher in psychopathy, Machiavellianism, and meaninglessness made more overt “utilitarian” judgments (i.e., fewer overt

“deontological” judgments). However, process dissociation clarified the underlying nature of this effect: when measured via the utilitarian parameter, the relations between all three individual difference variables and utilitarianism disappeared. Instead, all three personality variables correlated negatively with the deontology parameter. This finding suggests that people with antisocial personality traits do not experience greater utilitarian moral inclinations than other people; rather, they experience weaker deontological inclinations! Note that these findings were similar but weaker when we employed a regression analysis that assessed the effect of each personality trait controlling for the other traits, as well as controlling for social desirability and gender (in line with Bartels and Pizarro).

The findings of Study 1, Chapter 3 resolve theoretical confusion in the field: whereas previous work had characterized people with antisocial personality traits as “more utilitarian,” this does not fit with the dual-process model of moral judgments. Recall that utilitarianism is theorized to be driven primarily by cognitive evaluations of outcomes, whereas deontology is theorized to be driven primarily by affective reactions to harm. The antisocial personality traits examined here—particularly psychopathy—are characterized by equivalent levels of cognitive abilities compared to control participants, coupled with severe deficits in affect (Koenigs et al., 2007). Therefore, the dual-process model suggests that people with antisocial personality traits ought to experience less deontology but no less utilitarianism; this is exactly what we found. Process dissociation clarified the theoretical ambiguity of previous work by delineating the relation of each moral inclination to antisocial personality traits, further corroborating the dual-process model.

The final theoretical puzzle we examined was that of prosociality. Thus far, work on moral dilemmas has characterized utilitarianism as a willingness to cause harm in order to achieve the best overall outcome. While causing harm to achieve the best overall outcome does qualify as utilitarian motivation according to moral philosophers, it is an impoverished view of the nature of utilitarianism (Kagan, 1998; Mill, 1861/1998). True utilitarianism entails a focus on maximizing outcomes regardless of whether or not it is necessary to cause harm in order to achieve those outcomes. If psychologists are

genuinely tapping utilitarian moral inclinations, then utilitarianism ought to be associated with motivation to engage in prosocial behavior. To the author's knowledge, no work to date has examined whether this is, indeed, the case. Therefore, we resolved to examine the relations between deontology, utilitarianism, and prosocial motivation in Study 2, Chapter 3.

The results of Study 2, Chapter 3 indicated no significant relation between overt "utilitarian" (i.e., less overt "deontological") moral judgments on two of the three measures of prosocial motivation employed in the study. However, as we saw with the correlations between deontology, utilitarianism, and moral identity in Study 1, Chapter 2, null findings using overt judgments may be masking significant findings with the underlying parameters that do not shine through when deontology and utilitarianism are conflated. The one significant effect was a negative correlation between overt "utilitarian" (i.e., less "deontological") moral judgments and volunteerism. Taken at face value, this finding would suggest that the more participants made utilitarian judgments, the less motivated they were to volunteer. In other words, using overt judgments alone, utilitarianism does not look like an especially moral motivation; if anything, it looks like an amoral or even anti-moral motivation. Moreover, using overt judgments, there is no evidence that deontology entails prosocial motivation, unless one interprets the effect of volunteerism as reflecting greater motivation to volunteer among people who make more overt deontological judgments. Due to the non-independence error, this is only one of several possible interpretations.

Yet, the process dissociation results indicate a very different pattern of findings. When utilitarianism is measured via process dissociation, the negative relation with volunteerism disappears, becoming a null effect. Moreover, the correlation between the utilitarian parameter and prosocial decisions on the prosocial dilemmas was significant, suggesting that people who experience more utilitarian inclinations are more motivated to help others at a personal cost after all. This finding not only demonstrates how overt dilemma judgments can be misleading, it also enhances our understanding of utilitarianism. It suggests that utilitarianism is, in fact, not simply a motivation to harm others in order to achieve the best overall outcome—utilitarianism also entails a

willingness to make personal sacrifices to achieve the best overall outcome. This finding also accords with the finding from Study 1, Chapter 1, where the utilitarian parameter correlated positively with moral identity: both findings suggest there is something genuinely moral about utilitarianism. However, the utilitarian parameter did not significantly relate to prosocial motivations in the mundane helping scenarios or volunteerism. Thus, it may be that utilitarianism motivates some kinds of prosociality more than others.

When deontology is measured via process dissociation, the negative correlation between overt “utilitarian” judgments and volunteerism becomes clear: whereas volunteerism was not significantly correlated with the utilitarian parameter, it was positively correlated with the deontological parameter. This suggests that people who experienced stronger deontological inclinations also demonstrated stronger motivation to volunteer. Moreover, the deontology parameter also correlated positively with willingness to help in mundane helping scenarios. This finding corroborates the interpretation that deontology, as well as utilitarianism, motivates prosocial behavior. However, deontology did not correlate with prosocial decisions on the prosocial dilemmas. Thus, it may be that deontology motivates some kinds of prosociality more than others.

Finally, the prosocial dilemmas may be further broken down via factor analysis. Doing so reveals that not all prosocial dilemmas are the same; they cluster into three factors: *altruism dilemmas*, where actors must make a substantial, possibly fatal personal sacrifice in order to save one or more victims in grave peril, *honesty dilemmas*, where actors must decide whether to remain honest at a personal cost, or whether to act dishonestly for personal gain, and *charity dilemmas*, where actors must decide whether to donate money to a charity instead of spending it on the self. This analysis reveals interesting patterns: the relation between utilitarianism and prosocial motivation on prosocial dilemmas overall holds only for the altruism dilemmas: people scoring higher on the utilitarian parameter also demonstrated greater willingness to make a substantial personal sacrifice to save someone in dire straits; the deontology parameter was unrelated to decisions on altruism dilemmas. However, the deontology parameter correlated negatively with honesty dilemmas. This finding indicates that people scoring higher on the deontology

parameter were more likely to opt for honesty even when this imposes substantial costs on the self (e.g., failing to get hired). It is interesting that the utilitarian parameter failed to correlate with these dilemmas. This null effect suggests that utilitarianism is not about simply accepting the best outcome—when the “best outcome” benefits only the self, rather than attaining the best overall outcome for everyone in a situation, utilitarianism is no longer predictive. This further corroborates the view that utilitarianism is a genuinely moral motivation, rather than merely amoral outcome calculation. Finally, neither the deontology nor utilitarian parameters correlated significantly with charity dilemmas.

Considering the overall relations between deontology and utilitarianism with various prosocial measures, an interesting distinction emerges: deontology correlated positively with prosocial motivation in the mundane helping scenarios and the volunteerism scale; utilitarianism correlated positively with helping in the extreme altruism dilemmas. This pattern suggests not only that both parameters are tapping moral motivations, but that they are different kinds of moral motivations, or motivations that arise under different circumstances: deontology is predictive of moral motivation in mundane circumstances where help comes at a modest personal cost and will provide a modest benefit to the recipient, whereas utilitarianism is predictive of helping when there is a large personal cost and help will benefit recipients substantially. It remains unclear whether deontology and utilitarianism are tracking the relative costs to oneself, the relative benefits to others, or some other features of these different scenarios. However, this pattern suggests that people experience different prosocial motivations in different circumstances. Perhaps prosociality in mundane circumstances (e.g., doing a favor for a co-worker) is motivated by an affective reaction to the person’s need, whereas prosociality in extreme, life-or-death circumstances is motivated by consideration of the outcomes of action. This interpretation should be considered preliminary, but suggests an interesting distinction. Future work should examine whether this distinction is effective in motivating different kinds of actual prosocial behavior.

## 4.2 Implications

The current findings have implications for the dual-process model of moral judgments, the moral status of utilitarianism, and the interpretation of previous findings.

### 4.2.1 Implications: Dual-Process Model

The current findings have implications for the dual process model of moral judgments: they suggest that Greene and colleagues (2001) may be correct in their assertion that affective reactions to harm primarily drive judgments that harm is not acceptable, whereas cognitive deliberation primarily leads to judgments that harm is acceptable when it maximizes outcomes. Note that in the current studies, the utilitarian and deontological parameters were not, themselves, correlated (except a modest positive correlation in one case), yet each parameter correlated in expected directions with overt moral judgments. Moreover, the parameters may be manipulated independently, suggesting that one parameter mainly taps affective reactions to harm, whereas the other parameter mainly taps cognitive deliberation regarding outcomes—as predicted by the dual-process model. Indeed, these findings are inconsistent with single-process accounts (e.g., Kruglanski & Gigerenzer, 2011).

Recall that moral dilemma paradigm arose as a framework for integrating both cognition and emotion in moral judgments when other paradigms of the era argued that only one or the other was necessary and sufficient. Yet, as previous work was vulnerable to the dual-process error, it could not effectively rule out the possibility that a single process drove all dilemma judgments. Process dissociation helps overcome the non-independence error in previous dilemma research by measuring each parameter independently, and confirms that there appear to be at least two distinct processes in play when people make moral judgments. Therefore, given this data, it seems reasonable to continue research within the dual-process paradigm for the time being, rather than revert back to a debate regarding whether mostly visceral, affect-laden forces or mostly rational deliberation drives moral judgments.

### 4.2.2 Implications: Moral Status of Utilitarianism

These findings also have implications regarding the moral status of utilitarianism. Despite eloquent arguments for utilitarian ethics from moral philosophers (e.g., Mill, 1861/1998), the moral status of utilitarianism was in some doubt given recent evidence that utilitarian judgments are favored by people with antisocial personality traits (Bartels & Pizarro,



2011) and brain damage that impairs their social emotions (Koenigs et al., 2007). Note that we replicated these findings using overt dilemma judgments: people with antisocial personality traits made more utilitarian overt judgments (i.e., fewer deontological overt judgments). However, process dissociation analysis clarified the underlying relations: the utilitarian parameter did not significantly correlate with Machiavellianism, psychopathy, or meaninglessness. Rather, each of these individual difference variables correlated negatively with the deontology parameter. These findings confirm what Bartels and Pizarro, as well as Koenigs and colleagues, argued but could not determine in their own studies: the reason people scoring higher on antisocial traits favor utilitarian judgments is not because they experience *stronger* utilitarian inclinations; rather it is because they experience *weaker* deontological inclinations. This result is consistent with research on the nature of psychopathy and damage in ventromedial brain regions, given that such individuals demonstrate equivalent cognitive performance to control subjects despite profound emotional deficits (Saver & Damasio, 1991; Cleckley, 1955). This finding also helps resolve why people scoring lower on empathic concern also make more utilitarian overt moral judgments (Gleichgerricht, & Young, 2013). Finally, these results clarify that the utilitarian inclinations are not to be confused with a psychopathic perspective on human life. Mill (1861/1998) argued that utilitarianism implies a deep and abiding respect for human life, whereas psychopaths generally lack respect for human life (Cleckley, 1955).

Consistent with Mill's perspective that utilitarianism entails a moral position, the utilitarian parameter correlated positively with moral identity (internalization)—a measure of the degree to which morality forms part of the core self-concept (Aquino & Reed, 2002). Moral identity predicts a variety of prosocial outcomes such as increased charity donations, decreased prejudice, and decreased cheating (for a review, see Shao, Aquino, & Freeman, 2008). Intriguingly, the deontology parameter also correlated positively with moral identity, yet there was no significant correlation between moral identity and overt dilemma judgments. Thus, a process dissociation approach clarifies that utilitarianism may be a moral inclination after all, despite more overt utilitarian judgments among moral questionable individuals and a lack of relation between overt judgments and moral identity.

### 4.2.3 Implications: Interpreting Previous Results

Finally, these findings help resolve theoretical confusion regarding the nature of deontology and utilitarianism. Because overt, bi-polar moral judgments conflate judgments with the processes underlying judgments, interpreting them can be difficult. This was particularly true when researchers administered moral dilemmas to patients with brain damage in the VMPFC or people with antisocial personality traits (e.g., Koenigs et al., 2007, Bartels & Pizarro, 2011). When such participants make more ‘utilitarian’ judgments than other participants, what are we to make of this finding? Does it reflect powerful utilitarian inclinations, weak deontological ones, or some other possibility? Koenigs and colleagues titled their paper *Damage to the prefrontal cortex increases utilitarian moral judgments*. Is that the best way to characterize their results, given that increased utilitarian judgments are conflated with reduced deontological judgments?

Our findings suggest that increased utilitarian judgments results are due reduced deontological inclinations: people high on antisocial personality traits cared less about harm, and therefore experienced a weaker reaction to the thought of harm, which manifested as weaker deontological (but not utilitarian) inclinations. These findings are consistent with neuroimaging work demonstrating that people scoring higher in psychopathy demonstrated reduced activation in the amygdala while reading moral dilemmas (Glenn, Raine, & Schug, 2009). Therefore, perhaps Koenigs and colleagues should have titled their paper *Damage to the prefrontal cortex reduces deontological moral judgments*. To be fair, they and others have argued for this interpretation, but were unable to determine whether it was correct (Bartels & Pizarro, 2011; Greene, 2007a; Koenigs et al., 2007).

Moreover, the present results may clear up a misconception in the literature: the role of moral identity in moral judgments. Recently, Glenn, Koleva, Iyer, Graham, and Ditto (2010) measured psychopathy, moral identity, and moral judgments. They found the usual pattern: people scoring higher on psychopathy made more overt ‘utilitarian’ judgments. They also found, unsurprisingly, a negative correlation between psychopathy and moral identity (as moral identity is designed as a measure of the moral self-concept and predicts prosocial behavior, e.g., Aquino & Reed, 2002). Next, the researchers

examined whether moral identity mediated the relation between psychopathy and utilitarian judgment—but moral identity failed to correlate with utilitarian judgments. They argued “these results suggest that psychopathy is independently related to both a weaker moral identity as well as to more utilitarian moral judgment” (Glenn et al., 2010, p. 502).

There is another possible interpretation, however. Recall that there was no correlation between moral judgments and moral identity in Conway and Gawronski, Study 1, either. Yet, when measured via process dissociation, we found that moral identity was correlated with both the deontological and utilitarian parameters. This suggested that overt moral judgments obscure the true, positive relations between moral identity and both utilitarian and deontological inclinations. Moreover, recall in Conway, Bartels, and Pizarro, Study 4, we found that people scoring higher in psychopathy made more utilitarian judgments—an effect driven not by stronger utilitarian inclinations, but weaker deontological ones. Together, these findings suggest that moral identity might mediate the relation between psychopathy and the deontological parameter (and, thereby, moral judgments). This is an example of how process dissociation can provide improved methodological resolution to examine empirical claims, which ultimately has bearing on theoretical models of the nature of human moral psychology.

### 4.3 Limitations

The current work employed the now-common paradigm of providing participants with hypothetical scenarios and asking for judgments (some researchers ask for hypothetical behavioral responses; results tend to be similar, cf. Tassy, Oullier, Mancini, & Wicker, 2013). There are various disadvantages to such a paradigm: the dilemmas entail decisions between extreme actions (e.g., killing people; letting people die) in unlikely situations (e.g., time travel, paint bombs) of a hypothetical nature, wherein unrealistic actions occur (e.g., a single person’s falling body is capable of stopping a moving train). In short, moral dilemmas are absurd. The vagaries of moral dilemma research stem from its intellectual lineage as thought experiments in philosophy, in which realism is not a concern. Had psychologists invented the source material, rather than co-opted it from philosophy, perhaps researchers would ask participants to make decisions grounded in more realistic

scenarios. Given that this is not the case, the field is open to a number of criticisms. Each of these criticisms is valid, although they pertain to the entire moral dilemma paradigm rather than the current work exclusively. Nonetheless, it may be that none of the criticisms is strong enough to warrant disregarding the paradigm or empirical conclusions stemming from moral dilemma research.

First, critics have pointed out that participants may misjudge their actual reactions should they encounter circumstances like moral dilemmas in real life. This is almost certainly true: real life situations may involve far stronger visceral reactions than hypothetical scenarios given that harm should be considerably more salient in real life. Therefore, in real life people may be more likely to view harm as unacceptable. Indeed, a wealth of evidence supports this view: increasing the salience, vividness, or realism of harm tends to increase judgments that harm is not acceptable (Amit & Greene, 2012; Bartels, 2008), due to increased deontological inclinations evoked by visual imagery of the victim (Conway & Gawronski, 2013, Study 3). Moreover, when researchers attempted to increase the realism of moral dilemmas through such methods as employing virtual reality technology, participants who experienced stronger autonomic arousal made more judgments that harm is not acceptable (Navarette, McDonald, Mott, & Asher, 2011). Therefore, it may be that hypothetical scenarios underestimate the degree to which people are harm averse.

On the other hand, it may not be the case that the hypotheticality of moral dilemmas systematically underestimates deontology. Navarette and colleagues (2011) pointed out that in their virtual reality trolley dilemmas, the modal response to the trolley problem was still utilitarian. Therefore, even in realistic scenarios, if people can suppress their affective reactions, they may be able to act in ways that maximize the good even when doing so requires causing harm. Accordingly, police, soldiers, and emergency first responders commonly receive training on how to minimize emotional reactions in the face of traumatic events in order to keep a 'cool head' in dangerous and chaotic situations.

In any event, given that the field has concerned itself with responses evoked by different variants of dilemmas, a main effect of increased deontology in ‘real life’ settings does not endanger conclusions regarding the interaction between circumstances and deontology/utilitarianism—these conclusions remain valid, even if all decisions would shift closer to harm aversion in real-world situations. If emotionally evocative versions of dilemmas elicit stronger deontological but not utilitarian reasoning, whereas cognitive load impairs utilitarian but not deontological inclinations, then we are learning something about the structure of moral psychology.

It is also worth considering whether moral dilemmas are truly as hypothetical as they seem. O’Neill and Petrinovich (1998) pointed out,

“[Moral] dilemmas are not removed totally from the kinds of decisions policymakers sometimes have to make... In World War II, the British government had to decide whether or not to lure German bombers from highly populated London to less densely populated Kent, Surrey, or Sussex. The British government faced the following dilemma: Is it permissible to sacrifice a fewer number of individuals and save a greater number of individuals in the process? It was decided that it was morally impermissible to sacrifice fewer numbers of innocent people to save a greater number” (p. 362).

Closer to home, the Government of Manitoba recently faced the difficult decision of whether to intentionally divert the flooding Red River to destroy approximately 100 homes, or to allow the river to flood naturally, thereby destroying approximately 1000 homes. Ultimately, the government decided to embark on a program of intentional flooding (Canadian Broadcasting Corporation, 2011, May 13). Such a decision was parallel to the structure of moral dilemmas, even if the emotional stakes were lower (loss of property rather than life in this case).

Yet, there are many other cases that parallel moral dilemmas, ranging from triage of patients following a disaster, to assassination of political targets who cause strife (e.g., Osama bin Laden), to economic decisions (e.g., is it acceptable to fire some employees in order to save the company, thereby keeping the majority employed?). In fact, moral

dilemmas have been common in history: was it immoral for Ghengis Khan to destroy one city to facilitate the peaceful surrender of many others? Was it immoral for a dictator to impose Communism on the Russian people if it enabled them to industrialize and win the war? Was it immoral for America to drop the atomic bomb if it avoided many further casualties in combat? In each case, deontological and utilitarian positions disagree. Therefore, although moral dilemmas are themselves hypothetical, they parallel the structure of some real life situations. Therefore, by exploring the psychological factors that give rise to hypothetical moral judgments, we may gain insight into the factors that influence these real-life judgments.

If some real life dilemmas operate similarly to moral dilemmas, then people's answers to these real life questions may vary depending on (a) the degree of affect-laden reaction they experience, plus (b) the degree to which they deliberate regarding outcomes. Considering that victims often have stronger reactions to harm than perpetrators do (Baumeister, Stillwell, & Wotman, 1990), it may be that victims and perpetrators make systematically different moral judgments above and beyond self-serving bias (Epley & Dunning, 2001). This would be an interesting application of moral dilemma research. There is already some evidence for this perspective: third party observers who identify more with the victims than the perpetrators react more strongly to infractions, and have more difficulty forgiving transgressors (Brown, Wohl, & Exline, 2008). Hopefully research will continue in this vein.

A second challenge to moral dilemma research is that of *closed world assumptions*—assumptions contained within the dilemma itself that participants must assume to be true in order to answer the dilemma in a realistic manner (see Bennis, Medin, & Bartels, 2010). For example, the footbridge dilemma entails the assumption that pushing a person in front of a train will stop the train, and the torture dilemma entails the assumption that torture will result in effective actionable intelligence. If participants disagree with these assumptions, then the a priori logic of the dilemmas breaks down. For example, if participants believe that torture does not lead to accurate intelligence, then both deontology and utilitarianism should lead to judgments to avoid torture even when high explosives will kill people. Process dissociation improves the methodological resolution

for examining the processes leading the moral judgments, but it cannot resolve the issue regarding closed world assumptions, which is endemic to all moral dilemma research.

Third, there is a concern regarding the structure of moral dilemmas. Note that in all high-conflict/incongruent scenarios described in this paper and in other work, the utilitarian option entails acting in order to cause harm, whereas the deontological option entails refraining from action. Several studies have identified action/omission as one dimension that impacts responses on moral dilemmas (e.g., Petrinovich & O'Neill, 1996), but they changed the wording of the question rather than the nature of the action. Navarette and colleagues (2011) did reverse the nature of the action, but only within the trolley paradigm itself, which other theorists (e.g., Greene et al., 2001) argue does not pit strong deontological inclinations against utilitarian inclinations to the same degree that more emotionally evocative dilemmas do (e.g., footbridge dilemma). Moreover, all attempts to examine the role of action/omission suffer from the non-independence error, rendering it difficult to know which inclination is affected by action versus omission.

Therefore, in most dilemma research, action is conflated with utilitarianism and inaction with deontology, and in the few cases that disentangle these processes still conflate more judgments with the processes leading to judgments. This state of affairs risks conflating cautious or risk-averse response strategies with deontological inclinations, and action-oriented or risk-taking response strategies with utilitarian inclinations. It would be preferable if researchers counterbalanced a set of dilemmas in order to control for this confound, using process dissociation to determine which parameter (or whether both parameters) are affected by the action/omission distinction. In fact, Bertram Gawronski, Joel Armstrong, and I have resolved to do so. This is but one of several projects that have emerged out of the current project.

Finally, it is worthwhile to bear in mind that examination of the relations between the moral dilemma parameters and individual difference variables relied upon self-report scales to determine the nature of these individual differences, and self-report measures have well-known downsides: recollection of relevant events may be biased (Hawkins & Hastie, 1990), answers may shift due to contextual factors such as priming (Mussweiler,

& Strack, 1999), and systematic differences may occur depending on which particular measurement techniques are employed (e.g., Olson, Goffin, & Haynes, 2005). For example, Conway and Gawronski, Study 1 measured participants' moral identity using Aquino and Reed's (2002) moral identity measure. This measure presents participants with various moral terms, then asks them to indicate, for example, how important those traits are to themselves personally on a scale from *not true of me* to *completely true of me*. Another example: In Conway, Bartels, & Pizarro, Study 2, participants completed Devoe & Pfeffer's (2007) volunteerism scale, which presents several items asking participants to report how much they agree they would be willing to volunteer without pay on a scale from *strongly disagree* to *strongly agree*. Although such measures are often discussed as if they are directly tapping the construct of interest, this is not actually correct—rather, such measures are tapping *self-perceptions* of the degree to which one adheres to the construct of interest. In other words, the moral identity scale is really tapping self-perceptions of one's moral identity, which may be biased in ways that make people feel good about their moral status (Epley & Dunning, 2001). Similarly, the volunteerism scale is tapping self-perceptions of the degree to which one would be willing to volunteer, rather than actual volunteering behavior; these will be different to the degree that people are inaccurate regarding their predictions of future behavior. Therefore, the findings linking deontology and utilitarianism to various individual-difference measures should be interpreted with caution—they may reflect correspondence between these constructs and self-perceptions without necessarily mapping onto actual behavior.

That said, this criticism is limited to studies focused on the relation between the parameters and individual-difference variables; Conway and Gawronski Studies 2 and 3 experimentally manipulated theoretically-relevant variables (cognitive load and photos of victims) and demonstrated predicted, unique effects of these manipulations on the relevant parameters. These findings corroborate those examining individual-differences, thereby increasing confidence that the parameters related to more than mere self-perceptions. Finally, as discussed below, Peysakhovich, Conway, Greene, and Rand (in preparation) found that deontology uniquely predicted actual prosocial behavior using real money in economic games. This finding suggests that the predictive validity of the parameters extends beyond self-report scales to actual behavior. Future work will



hopefully continue to incorporate real behavioral measures in order to further increase confidence in the validity of the parameters.

## 4.4 Future Directions

Now that the present work has validated a process dissociation measure of deontological and utilitarian inclinations, this technique may be applied to a host of theoretical questions in the field.

### 4.4.1 Gender Differences in Moral Judgments

One of the questions process dissociation may shed light on is gender differences in moral judgments. Bartels and Pizarro (2011) found that male participants made more overt utilitarian judgments than female participants overall. Fumagalli et al. (2010) also found a gender difference in overt judgments, which they isolated to personal dilemmas only (those involving direct harm to the victim); they found no gender differences on impersonal dilemmas (those where harm is mechanically mediated) or non-moral dilemmas. However, as they measured overt judgments only, these findings are vulnerable to the non-independence error. Therefore, they might reflect increased utilitarian inclinations among men, increased deontological inclinations among women, or some combination thereof.

Consistent with Gilligan's (1982) distinction between women's preference for an ethic of care as distinct from men's preference for an ethic of justice, Fumagalli and colleagues interpreted the gender difference in moral judgments as reflecting both processes:

“Female moral reasoning seems directed to avoid harming other people, to place high value on social relationships and to fulfill other individuals' expectations. Conversely, male moral thought hinges on the abstract principles of justice and fairness and on an individualistic stance” (2010, p. 223).

Yet, due to the non-independence error, techniques examining overt judgments cannot determine whether Fumagalli and colleagues' view is the best interpretation, or whether other interpretations better reflect the gender difference. Conversely, process dissociation

can distinguish whether gender differences result from increased utilitarian processing in men or reduced utilitarian processing in women, or increased deontological processing in women or decreased deontological processing in men—or some combination thereof.

Indeed, Conway, Friesdorf, and Gawronski (in preparation) have conducted a mega-analysis on all available process dissociation data sets (39 datasets, total  $N = 5990$ ). Note that at this sample size, all effects are significant, but significance testing is no longer informative, so we rely instead on analysis of effect size. Overall, we replicated the finding that men make more overt utilitarian judgments than women, Cohen's  $d = -.54$ . However, a process dissociation analysis revealed this effect to be driven largely by women scoring higher than men on the deontology parameter, Cohen's  $d = .59$ . Although men also scored higher than women on the utilitarian parameter, this effect was considerably smaller, Cohen's  $d = -.10$ . Therefore, these findings provide improved support for Gilligan's (1982) contention that women experience stronger affective reactions to harm (an element of the ethic of care). These findings also support Gilligan's contention that men engage in more dispassionate, pragmatic moral reasoning (an ethic of justice) than women. Yet, the difference between men and women in terms of utilitarianism is smaller than the difference in terms of deontology, suggesting that much of the variance in overt judgments noted by Bartels and Pizarro (2011) and Fumagalli et al. (2010) was driven by gender differences in deontology. That said, these findings also suggest that there is considerable within-gender variability on both deontology and utilitarianism.

#### 4.4.2 The Role of Threat in Moral Judgments

The current work, as well as the rest of the moral dilemma literature, corroborates the dual-process model suggesting that deontological judgments are driven primarily by negative affect resulting from an empathic response to the harm suffered by victims, whereas utilitarian judgments are driven primarily by cognitive evaluations of outcomes. Yet, researchers have assumed, rather than tested, whether the negative affect involved in deontological responses is truly other-directed. Granted, a wealth of findings (including Conway & Gawronski, Study 3) suggest that the imagining harm in more vivid fashion increases deontological judgments, consistent with the idea that making the victims' pain

more salient increases deontology. This suggests that the affect involved in deontology might be other-directed.

Yet, consider that moral dilemmas essentially invite people to become ‘butchers’ in the service of the greater good. The more vividly one imagines causing violent harm to another (albeit with good intentions), the more one may experience negative affect due to implications *for the self*. After all, people view harm they cause directly as worse than other kinds of harm (Cushman, Young, & Hauser, 2006) and experience moral failings as threatening to the self (Blasi, 1980; Cameron & Payne, 2012). Therefore, imagining causing harm (i.e., making a utilitarian judgment) might threaten people’s self-concept, eliciting negative emotions that motivate them to make a deontological judgment.

If considering harm elicits negative affect due to self-threat, deontological judgments may increase either because threat increases deontological motivations or reduces utilitarian motivation. Whereas the first finding would corroborate the dual-process model—which posits the negative affect elicited by harm affects only deontological processing—the second finding would require a re-evaluation of the dual-process model, which posits that only deontology is strongly motivated by affect. I have planned five studies to investigate the role of threat to the self in moral dilemma judgments. Studies 1 and 2 will prime self vs. other focus vs. control either explicitly (via a mirror vs. photos of victims, Phillips & Silvia, 2005) or implicitly (via subliminal priming of participants’ own name vs. others’ names, Silvia, 2012) as participants complete a moral dilemma battery (Conway & Gawronski, 2013). I expect to replicate my previous finding that the other-focus condition increases deontological judgments compared to the control condition—an effect driven by the deontological (but not utilitarian) process dissociation parameter. If deontological judgments are also driven by threat to the self, then deontological judgments ought to be higher in the self-focus condition as well—though it remains to be seen whether this difference will manifest on the deontological or utilitarian parameter when measured via process dissociation.

If deontological inclinations are driven by threat to the self elicited by imagining causing harm, then reducing perceptions of self-threat via physical cleansing (Study 3) or self-

affirmation (Study 4) should reduce deontological judgments. Cleansing oneself reduces shame and guilt produced by moral failings (Zhong & Liljenquist, 2006), so if causing harm is threatening to the self, then participants who have an opportunity to cleanse themselves with a hand wipe might make fewer deontological judgments compared to those who merely get to view the hand wipe. Study 4 will employ a self-affirmation manipulation known to reduce defensiveness to threatening information (Sherman, Nelson, & Steele, 2000) by increasing the salience of values that are central to a person's self-image (Steele, 1988). If self-affirmation reduces self-threat of harming someone, this manipulation should have a similar effect as cleansing in Study 3.

Finally, Study 5 will examine whether threatening the self will increase deontological inclinations. Previous research has established that recalling previous immoral behavior threatens the moral self, which motivates an increase in moral striving to prove moral worth (Sachdeva, Iliev, & Medin, 2009). These effects are limited, however, to the recall of recent, concrete behaviors (Conway & Peetz, 2012). If deontological moral judgments are driven purely by empathic reactions to suffering, then one's previous behavior ought to have no impact on judgments, as it is irrelevant to the degree of suffering experienced by the victim. Conversely, if deontological moral judgments are motivated by threat to the moral self, then the prior status of the moral self ought to influence the strength of deontological moral judgments. To examine this possibility, we will manipulate moral self-perceptions by asking participants to recall recent moral or immoral behavior (see Conway & Peetz, 2012) before they complete the moral dilemma battery. Together, these five studies should shed light on the role of threat to the self in moral judgments.

#### 4.4.3 Other Process Dissociation Projects

In addition to clarifying gender differences and the role of self-threat in moral judgments, process dissociation has proven useful in clarifying the nature of other moral processing. For example, Wisneski, Conway, and Skitka (in preparation) found that deontology parameter correlated positively with a measure of individual differences in the tendency to moralize political positions (e.g., The Iraq War, Abortions). This suggests that people who are more likely to consider the moral implications of political policies also tend to viscerally react to the thought of harm. Conversely, there was a negative correlation

between the utilitarian parameter and individual differences in the tendency to moralize everyday life (e.g., lying to a friend, impure behavior). This suggests that people who are more likely to consider the moral implications of everyday behavior are also less likely to accept pragmatic moral trade-offs; perhaps they prefer to stick with moral principles or with their intuitions.

Consistent with these considerations, Wisneski, Conway, and Skitka found the deontology parameter positively correlated with the individualizing foundations (harm and fairness) but not the binding foundations (respect for authority, loyalty, and purity) of the moral foundations questionnaire (Graham, Haidt, & Nosek, 2009). Conversely, utilitarianism was unrelated to any of the moral foundations except negatively related to purity. This suggests that people who care relatively more about harm and fairness tend to experience stronger inclinations to avoid causing harm. On the other hand, people who care relatively more about purity (e.g., avoiding drugs and sexual perversion) are most skeptical of utilitarian tradeoffs. These findings suggest that individual differences in moral inclinations are not relevant only in the artificial circumstances of moral dilemmas, but may relate to broader views, such as the moralization of politics or everyday life.

In another project, Peysakhovich, Conway, Greene, and Rand (in preparation) used individual differences in deontological and utilitarian inclinations to predict prosocial responses on economic games. Participants completed a dictator game, where they received a bonus (\$3) and could give any portion of it to an anonymous person, who would have no control over the situation. Participants also completed the public goods game, where they were given a bonus (\$3) and could donate any portion of that bonus to a communal pot, which would be doubled then shared by four players. Therefore, donations were beneficial for group outcomes, but detrimental for one's own outcomes. We recorded the amount participants donated across both games.

Peysakhovich, Conway, Greene, and Rand found that the deontology parameter predicted cooperation on both types of economic games, whereas utilitarianism did not. These results held above and beyond gender and other control variables. This may be viewed as a conceptual replication of Conway, Bartels, and Pizarro (under review) Study 2, where

deontology predicted prosocial behavior in mundane contexts, such as offering ordinary assistance to another person. It may be that people who experience stronger affective reactions to harm are also more likely to experience empathy for others in need under ordinary circumstances, leading to more prosocial responses. In Conway, Bartels, and Pizarro, Study 2, utilitarianism also predicted prosocial responses in extreme circumstances. Therefore, perhaps if Peysakhovich and colleagues ran an extreme version of the dictator game where people had absolute control over whether another person lives or dies, then utilitarianism might predict prosocial responses as well. Perhaps an experimentally-appropriate analogue of this situation may be created, allowing for an examination of this possibility.

Another interesting study to consider would be a replication and extension of Conway and Gawronski, Study 3. In that study, participants who read dilemmas accompanied by a photo of the individual victim of harm for each dilemma demonstrated increased deontological, but not utilitarian, inclinations, compared to a control group of participants who read dilemmas without photos. This finding fits with the dual process model, as it suggests that a focus on the single victim enhanced the salience of harm potentially dealt to that individual, which increased inclinations to avoid causing harm. However, the question remains what would occur if similar photos would be presented for other people in each dilemma, those who are about to suffer impending harm unless harm is instead dealt to the individual target. The dual-process model suggests that judgments favoring the five original victims are driven by cognitive evaluations of overall outcomes, whereas judgments favoring the single victim are driven by an affective reaction against causing harm. Would showing photos of the other victims enhance the salience of the harm they are about to suffer? Presumably it would. The question then is whether this process would lead to increased utilitarianism, decreased deontology, or perhaps some other pattern, such as both increased utilitarianism and deontology. If enhancing the salience of harm to the existing victims increases utilitarianism (without impacting deontology), this finding would provide the first evidence of a link between affect and utilitarianism—thereby suggesting a revision to the existing dual-process model. Therefore, such a study might be very useful in further clarifying the nature of utilitarianism and deontology.

Finally, Bernhard, Conway, and Greene (in preparation) found a link between the deontology and utilitarian parameters and genetic variation in a gene related to oxytocin, a hormone involved in a wide variety of social behaviors (for a review, see Lee et al., 2009). This gene polymorphism (identified as rs237889) has three alleles: dominant G/G homozygous, mixed A/G heterozygous, and recessive A/A homozygous. The variants are linked to differential ability to detect the hormone oxytocin in the nervous system, a hormone involved in many social behaviors, such as maternal care (Bosch & Neumann, 2012) and cooperation in economic games (Israel et al., 2009).

Bernhard, Conway, and Greene found an interaction between parameter and genotype, such that that people with A/A alleles scored higher on deontology than people with A/G alleles, who, in turn, scored higher on deontology than people with G/G alleles.

Utilitarianism showed the opposite pattern: people with G/G alleles scored higher on utilitarianism than people with A/G alleles, who, in turn, scored higher on utilitarianism than people with A/A alleles. These results held above and beyond gender. These findings suggest that individual differences in deontological and utilitarian inclinations may be related to individual differences in genetic make-up. Moreover, the genetic and moral judgment perspectives mesh nicely: people with G/G allele genotypes are more prone to antisocial behavior (Pluijm, 2012), and also experience the lowest degree of deontological inclinations, reminiscent of people with antisocial personality traits in Conway, Bartels, and Pizarro, (under review) Study 1. Obviously, there is much more work to be done to clarify the genetic correlates of moral inclinations, but this study is a first step in this new direction.

## 4.5 Summary

In sum, this project expands and refines the study of human moral psychology by providing an important methodological and theoretical advance. One product of this work is a new tool to enable researchers to examine more fine-grained questions regarding the relations between their variables of interest and utilitarianism and deontology. Another outcome is increased insight into the nature of the constructs in play. The application of process dissociation to moral judgments has (a) revealed obscured relations between moral identity and deontological and utilitarian inclinations, (b) clarified that the two

inclinations are, indeed, statistically independent, with different contextual and correlational profiles corresponding to theoretical predictions, (c) clarified that the relation between moral judgment and antisocial personality traits are driven by reduced deontological inclinations, rather than increased utilitarian ones, and (d) indicated that both utilitarianism and deontology predict some elements of prosociality, consistent with philosophical perspectives. Therefore, the current work provides a methodological and theoretical advance that pushes the current state of knowledge regarding human moral psychology one step further.



## References

- Aquino, K., & Reed, A. II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83, 1423-1440. doi:10.1037//0022-3514.83.6.1423
- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, 121, 154-161. doi:10.1016/j.cognition.2011.05.010
- Baumeister, R.F., Stillwell, A., & Wotman, S. R. (1990). Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology*, 59, 994-1005.
- Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, 5, 187-202. doi:10.1177/1745691610362354
- Bernhard, R., Conway, P., & Greene, J. (in preparation). Deontology and utilitarianism relate to the AA and GG homozygous alleles, respectively, of the oxytocin reception gene polymorphism rs237889.
- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, 88, 1-45. doi:10.1037/0033-2909.88.1.1
- Bosch, O. J., & Neumann, I. D., (2012). Both oxytocin and vasopressin are mediators of maternal care and aggression in rodents: From central release to sites of action. *Hormones and Behavior*, 61, 293-303. doi:10.1016/S0079-6123(08)00422-
- Brown, R. P., Wohl, M. J. A., & Exline, J. J. (2008). Taking up offenses: Second-hand forgiveness and identification with targets of transgressions. *Personality and Social Psychology Bulletin*, 34, 1406-1419.
- Cameron, C. D., & Payne, B. K. (2012). Regulating compassion influences the moral concept. *Psychological Science*, 23, 225-229. doi:10.1177/0956797611430334
- Canadian Broadcasting Corporation. (2011, May 13). *Manitoba flood diversion on for Saturday: premier*. Retrieved from <http://www.cbc.ca/news/canada/story/2011/05/13/mb-flooding-lake-manitoba.html>
- Conway, P., Friesdorf, R., & Gawronski, B. (in prep). Gilligan in trolleyland: Resurrecting the debate over gender differences in morality.

- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, 104, 216-235. doi:10.1037/a0031021
- Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, 38, 907-919.
- Devoe, S. E., & Pfeffer, J. (2007). Hourly payment and volunteering: The effect of organizational practices on decisions about time use. *Academy of Management Journal*, 50, 783–798.
- Cleckley, H. (1955). *The mask of sanity*. St. Louis: C. V. Mosby.
- Cushman, F., Young, L., & Hauser, M., (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science*, 17, 1082-1089.
- Epley, N., & Dunning, D. (2000). Feeling “Holier than thou”: Are self-serving assessments produced by errors in self or social prediction? *Journal of Personality and Social Psychology*, 79, 861-875.
- Fischer, J.M., and Ravizza, M. (1992). Introduction. In J.M. Fischer and M. Ravizza (Eds.). *Ethics: Problems and Principles*, 1–57. Orlando: Harcourt Brace Jovanovich.
- Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Spota, S., Zago, S., Lucchiari, C., Consonni, D., Nordio, F., Pravettoni, G., Cappa, S. & Priori, A. (2010). Gender related differences in moral judgments. *Cognitive Processing*, 11, 219-226. doi:10.1007/s10339-009-0335-2
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS ONE*, 8, e60418. doi:10.1371/journal.pone.0060418
- Glenn, A. L., Raine, A., & Schug, R. A. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry*, 14, 5-9.
- Glenn, A.L., Koleva, S., Iyer, R., Graham, J., Ditto, P.H. (2010). Moral identity in psychopathy. *Judgment and Decision Making*, 5, 497-505.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029-1046.

- Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*, 322-323. doi:10.1016/j.tics.2007.06.004
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144-1154. doi:10.1016/j.cognition.2007.11.004
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105-2108.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814-834. doi:10.1037//0033-295X.108.4.814
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311-327. doi:10.1037/0033-2909.107.3.311
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.
- Israel, S., Lerer, E., Shalev, I., Uzevovsky, F., Riebold, M., Laiba, E., Bachner-Melman, R., Maril, A., Bornstein, G., Knafo, A., & Ebstein, R. P. (2009). The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task. *PLoS One*, *4*, e5535. doi:10.1371/journal.pone.0005535.
- Kagan, S. (1998). *Normative Ethics*. Westview Press: Boulder, CO.
- Kant, I. (1785/1959). *Foundation of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* *446*, 908-911. doi:10.1038/nature05631
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 347-480). Chicago: Rand McNally.
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgments are based on common principles. *Psychological Review*, *118*, 97-109. doi:10.1037/a0020762

- Lee, H., Macbeth, A. H., Pagani, J. H., & Young the 3rd, W. S. (2009). Oxytocin: The great facilitator of life. *Progress in Neurobiology*, 88, 127-151. doi:10.1016/j.pneurobio.2009.04.001.
- Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology*, 18, 193-197.
- Mill, J. S. (1861/1998). *Utilitarianism*. In R. Crisp (Ed.), New York: Oxford University Press.
- Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35, 136-164.
- Navarette, C. D., McDonald, M. M., Mott, M. L. & Asher, B. (2011). Morality: Emotion and action in a simulated three-dimensional “trolley problem.” *Emotion*, 12, 364–370.
- Olson, J. M., Goffin, R. D., & Haynes, G. A. (2007). Relative versus absolute measures of attitudes: Implications for predicting diverse attitude-relevant criteria. *Journal of Personality and Social Psychology*, 93, 907-925.
- Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, 20, 272-314. doi:10.1080/10463280903162177
- Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology & Sociobiology*, 17, 145-171.
- Peysakhovich, A., Conway, P., Greene, J., & Rand, D. (in preparation). The relation between moral dilemmas and cooperation in economic games.
- Pluijm, S. v. d. (2012). De invloed van oxytocine- en serotonine genen op informatieverwerkingsvaardigheden die een rol spelen bij de sociale interactie. Retrieved from: <http://irs.ub.rug.nl/dbi/50d44b2e77c18>
- Phillips, A. G., & Silvia, P. J. (2005). Self-awareness and the emotional consequences of self-discrepancies. *Personality and Social Psychology Bulletin*, 31, 703–713. doi:10.1177/0146167204271559
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, 20, 523-528. doi:10.1111/j.14679280.2009.02326.x
- Saver, J. L. & Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, 29, 1241–1249. doi:10.1016/0028-3932(91)90037-9

- Shao, D., Aquino, K., & Freeman, D. (2008). Beyond moral reasoning: A review of moral identity research and its implications for business ethics. *Business Ethics Quarterly*, 18, 513-540.
- Sherman, D. A. K., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self: Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin*, 26, 1046–1058. doi:10.1177/01461672002611003
- Silvia, P. (2012). Mirrors, masks, and motivation: Implicit and explicit self-focused attention influence effort-related cardiovascular reactivity. *Biological psychology*, 90, 192-201. doi:10.1016/j.biopsycho.2012.03.017
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261-302). New York: Academic Press.
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119, 454-458. doi:10.1016/j.cognition.2011.01.018
- Tassy, S. Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, 4, .e00250. doi:10.3389/fpsyg.2013.00250
- Wisneski, D., Conway, P., & Skitka L. (in preparation). The moralization of politics versus everyday life: A test of moral foundations theory.
- Zhong, C., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, 313, 1451–1452.

## Appendix 4A: Copyright Permission



AMERICAN  
PSYCHOLOGICAL  
ASSOCIATION

Paul Conway  
University of Western Ontario

INVOICE NO. N/A  
Federal Tax I.D. 53-0205890  
Date: February 27, 2013

IN MAKING PAYMENT REFER TO THE ABOVE INVOICE NUMBER

APA Permissions Office

IF THE TERMS STATED BELOW ARE ACCEPTABLE, PLEASE SIGN AND RETURN ONE COPY TO APA. RETAIN ONE COPY FOR YOUR RECORDS. PLEASE NOTE THAT PERMISSION IS NOT OFFICIAL UNTIL APA RECEIVES THE COUNTERSIGNED FORM AND ANY APPLICABLE FEES.

This Agreement replaces the Agreement dated January 10, 2012.

Revision: Year changed, citation changed

Request is for the following APA-copyrighted material:

- Deontological and utilitarian inclinations in moral decision making: A process dissociation approach.  
Conway, Paul; Gawronski, Bertram. Journal of Personality and Social Psychology, Vol 104(2), Feb 2013, 216-235. doi: 10.1037/a0031021

For the following use: **Non-Commercial Research or Educational Use in:** a) the print version of the final dissertation document; b) the digital version of the final dissertation document posted online provided all conditions in item (3) below are met.

File: Conway, Paul (author)

Permission is granted for the nonexclusive use of APA-copyrighted material specified above contingent upon fulfillment of the conditions indicated below:

1. The fee is waived.
2. The reproduced material must include the following credit line: Copyright © [indicate year] by the American Psychological Association. Reproduced [or Adapted] with permission. The official citation that should be used in referencing this material is [list the original APA bibliographic citation]. No further reproduction or distribution is permitted without written permission from the American Psychological Association.
3. **For online use:**
  - (a) The credit line must appear on the first screen on which the APA content appears.
  - (b) The online posting is permitted only on a non-commercial, secure and restricted web site. If condition (b) cannot be satisfied then the abstract may be reproduced along with a link to the journal table of contents at <http://psycnet.apa.org/journals/psp/>.

This agreement constitutes permission to reproduce only for the purposes specified on the attached request and does not extend to future editions or revisions, derivative works, translations, adaptations, promotional material, or any other formats or media. Permission applies solely to publication and distribution in the English language throughout the world, unless otherwise stated. No changes, additions, or deletions to the material other than any authorized in this correspondence shall be made without prior written consent by APA.

This permission does not include permission to use any copyrighted matter obtained by APA or the author(s) from other sources that may be incorporated in the material. It is the responsibility of the applicant to obtain permission from such other sources.

ACCEPTED AND AGREED TO BY:

\_\_\_\_\_  
Applicant  
Date Feb 28, 2013

PERMISSION GRANTED ON ABOVE TERMS:

\_\_\_\_\_  
for the American Psychological Association  
Date February 27, 2013

## Appendix 4B: Ethics Approval

Conway and Gawronski (2013) Study 1



**Department of Psychology** The University of Western Ontario

### Use of Human Subjects - Ethics Approval Notice

<b>Review Number</b>	08 09 06	<b>Approval Date</b>	08 09 08
<b>Principal Investigator</b>	Bertram Gawronski/Paul Conway	<b>End Date</b>	09 04 30
<b>Protocol Title</b>	How do we make moral judgments		
<b>Sponsor</b>	n/a		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: <http://www.uwo.ca/research/ethics/>)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:

- a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
- b) all adverse and unexpected experiences or events that are both serious and unexpected;
- c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2087-2009 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

*This is an official document. Please retain the original in your files*

## Conway and Gawronski (2013) Study 2



Department of Psychology The University of Western Ontario

## Use of Human Subjects - Ethics Approval Notice

<b>Review Number</b>	<b>08 10 24</b>	<b>Approval Date</b>	<b>08 10 28</b>
<b>Principal Investigator</b>	<b>Bertram Gawronski/Paul Conway</b>	<b>End Date</b>	<b>09 04 30</b>
<b>Protocol Title</b>	<b>Does memory influence moral and evaluative judgments?</b>		
<b>Sponsor</b>	<b>n/a</b>		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: <http://www.uwo.ca/research/ethics/>)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:

- a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
- b) all adverse and unexpected experiences or events that are both serious and unexpected;
- c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

Clive Seligman Ph.D.

Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2087-2009 PREB are: David Dozois, Bill Fisher, Riley Hinson and Steve Lupker

CC: UWO Office of Research Ethics

*This is an official document. Please retain the original in your files*



## Conway and Gawronski (2013) Study 3



Department of Psychology The University of Western Ontario

## Use of Human Subjects - Ethics Approval Notice

<b>Review Number</b>	<b>11 09 42</b>	<b>Approval Date</b>	<b>11 09 29</b>
<b>Principal Investigator</b>	<b>Jim Olson/Paul Conway</b>	<b>End Date</b>	<b>12 04 30</b>
<b>Protocol Title</b>	<b>Making decisions, viewing video clips, rating products, and judging online media</b>		
<b>Sponsor</b>	<b>n/a</b>		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: <http://www.uwo.ca/research/ethics/>)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:

- a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
- b) all adverse and unexpected experiences or events that are both serious and unexpected;
- c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

Clive Seligman Ph.D.

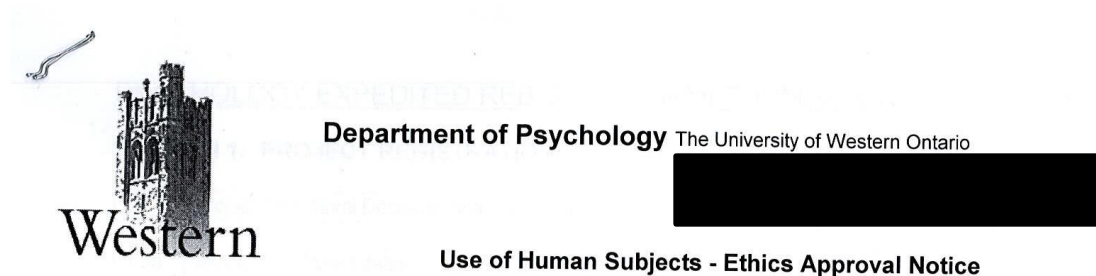
Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2011-2012 PREB are: Mike Atkinson (Introductory Psychology Coordinator), Rick Goffin, Riley Hinson, Albert Katz (Department Chair), Steve Lupker, and TBA (Graduate Student Representative)

CC: UWO Office of Research Ethics

*This is an official document. Please retain the original in your files*

## Conway, Bartels, &amp; Pizarro (under review) Study 1



<b>Review Number</b>	<b>12 02 18</b>	<b>Approval Date</b>	<b>12 02 17</b>
<b>Principal Investigator</b>	<b>Jim Olson/Paul Conway</b>	<b>End Date</b>	<b>12 04 30</b>
<b>Protocol Title</b>	<b>Moral decisions and personality style</b>		
<b>Sponsor</b>	<b>n/a</b>		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: <http://www.uwo.ca/research/ethics/>)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:

- a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
- b) all adverse and unexpected experiences or events that are both serious and unexpected;
- c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the PREB.

Clive Seligman Ph.D.

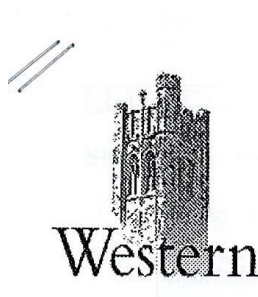
Chair, Psychology Expedited Research Ethics Board (PREB)

The other members of the 2011-2012 PREB are: Mike Atkinson (Introductory Psychology Coordinator), Rick Goffin, Riley Hinson, Albert Katz (Department Chair), Steve Lupker, and Karen Dickson (Graduate Student Representative)

CC: UWO Office of Research Ethics

*This is an official document. Please retain the original in your files*

## Conway, Bartels, &amp; Pizarro (under review) Study 2



Department of Psychology      The University of Western Ontario

**Use of Human Subjects - Ethics Approval Notice**

<b>Review Number</b>	<b>12 05 05</b>	<b>Approval Date</b>	<b>12 05 07</b>
<b>Principal Investigator</b>	<b>OLSON, J.</b>	<b>End Date</b>	<b>12 12 20</b>
<b>Protocol Title</b>	<b>JUDGING DECISIONS AND BEHAVIOR</b>		
<b>Sponsor</b>	<b>n/a</b>		

This is to notify you that The University of Western Ontario Department of Psychology Research Ethics Board (PREB) has granted expedited ethics approval to the above named research study on the date noted above.

The PREB is a sub-REB of The University of Western Ontario's Research Ethics Board for Non-Medical Research Involving Human Subjects (NMREB) which is organized and operates according to the Tri-Council Policy Statement and the applicable laws and regulations of Ontario. (See Office of Research Ethics web site: <http://www.uwo.ca/research/ethics/>)

This approval shall remain valid until end date noted above assuming timely and acceptable responses to the University's periodic requests for surveillance and monitoring information.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the PREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of research assistant, telephone number etc). Subjects must receive a copy of the information/consent documentation.

Investigators must promptly also report to the PREB:

- a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
- b) all adverse and unexpected experiences or events that are both serious and unexpected;
- c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to the PREB for approval.

Members of the PREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to the study. [Redacted]

Riley Hinson, Ph.D.

Acting Chair (in the absence of Clive Seligman), Psychology Expedited Research Ethics Board (PREB)

The other members of the 2011-2012 PREB are: Mike Atkinson (Introductory Psychology Coordinator), Rick Goffin, Riley Hinson, Albert Katz (Department Chair), Steve Lupker, and Karen Dickson (Graduate Student Representative)

CC: UWO Office of Research Ethics

*This is an official document. Please retain the original in your files*

## Curriculum Vitae

**PAUL CONWAY**

### EDUCATION

- Post-Doctoral Fellow**, 2013-2015, University of Cologne  
 Supervisor: Dr. Thomas Mussweiler
- 2013 Ph.D. Psychology**, The University of Western Ontario  
 Supervisor: Dr. James Olson  
 Co-supervisor: Bertram Gawronski
- 2008 M.Sc. Psychology**, The University of Western Ontario  
 Supervisor: Dr. James Olson
- 2005 B.A. Psychology (Honors)**, Simon Fraser University  
 Supervisor: Dr. Dennis Krebs

### AWARDS & HONOURS

- Student Publication Award 2012, Society for Personality and Social Psychology
- Student Poster Award 2012, Society for Personality and Social Psychology
- Honorable Mention, Outstanding Research Award 2012, Society for Personality and Social Psychology
- Runner-Up, Social Science Oral Presentation 2010, Western Research Forum
- Runner-Up, Student Poster Award 2008, Society for Personality and Social Psychology
- Nominated, Teaching Assistant Award 2007, 2008, 2009, 2010, 2012, The University of Western Ontario

### PUBLICATIONS

- Conway, P.**, & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision-making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*, 216-235. doi:10.1037/a0031021
- Hafer, C. L., **Conway, P.**, Cheung, I., Malyk, D., & Olson, J. M. (2012). The relation between identification with a target and the perceived importance of justice. *Basic and Applied Social Psychology*, *34*, 395-409. doi:10.1080/01973533.2012.711693.
- Conway, P.**, & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, *38*, 907-919. doi:10.1177/0146167212442394
- Olson, J. M., Cheung, I., **Conway, P.**, Hutchison, J., & Hafer, C. L. (2011). Distinguishing two meanings of moral exclusion: Irrelevance of fairness vs. rationalized harm-doing. *Social Justice Research*, *24*, 1-26.

Olson, J. M., Hafer, C. L., Cheung, I., & **Conway, P.** (2009). Deservingness, the scope of justice, and actions toward others. In A. C. Kay, D. R. Bobocel, M. P. Zanna, & J. M. Olson (Eds.), *The psychology of justice and legitimacy: The Ontario symposium* (Vol. 11, pp. 125-149). New York: Psychology Press.

## MANUSCRIPTS UNDER REVIEW

**Conway, P.**, Bartels, D. M., & Pizarro, D. A. (under review). Are psychopaths really more utilitarian? Process dissociation clarifies the relations between utilitarianism, psychopathy, and prosociality. *Psychological Science*.

Förster, G., **Conway, P.**, & Strack, F. (under review). Knowledge activation or response substitution? Specific-general effects in judgments of well-being. *European Journal of Social Psychology*.

Maxwell-Smith, M., Seligman, C., **Conway, P.**, & Cheung, I. (under review). The amplification effect of individual differences in commitment to beliefs on perceived belief dissimilarity. *Journal of Personality*.

## FUNDING

\$8,000, Graduate Research Scholarship, 2011, The University of Western Ontario

\$5,900, Canada Graduate Scholarship: Michael Smith Foreign Study Supplement, 2009, Social Sciences and Humanities Research Council of Canada

\$105,000, Joseph-Armand Bombardier Canada Graduate Scholarship, 2008-2011, Social Sciences and Humanities Research Council of Canada

\$15,000, Ontario Graduate Scholarship, 2008, Government of Ontario (Declined)

\$8,000, Graduate Research Scholarship, 2007, The University of Western Ontario

\$17,500, Canada Graduate Scholarship: Masters, 2006-2007, Social Sciences and Humanities Research Council of Canada

\$4000, Graduate Research Scholarship, 2006, The University of Western Ontario

\$4000, Summit Entrance Scholarship, 2000, Simon Fraser University

## RESEARCH EXPERIENCE

September 2013-2015, Mussweiler Social Cognition Lab, University of Cologne

May-August 2010, Strack Social Cognition Lab, University of Würzburg

September 2007-2013, Gawronski Social Cognition Lab, The University of Western Ontario

September 2006-2013, Olson Attitudes & Justice Lab, The University of Western Ontario

January 2004-June 2005, McFarland Social Cognition Lab, Simon Fraser University

September 2003-May 2004, Krebs Morality Lab, Simon Fraser University

**TEACHING EXPERIENCE**

Summer 2013, Introduction to Social Psychology, Western  
Winter 2013, The Psychology of Persuasion, Western  
Fall 2012, Child Development, King's University College  
2012-2013, Introduction to Developmental Psychology, King's University College  
Summer 2012, Introduction to Social Psychology, Western  
Summer 2011, Introduction to Social Psychology, Western (Co-instructor)

**SERVICE**

2012-2014, President and Past-President, Graduate Student Council, Society for  
Personality and Social Psychology  
2011-2012, Graduate Representative, Western Psychology Appointments Committee  
2011-2012, Brownbag Speaker Series Coordinator, Western Social Area  
2010-2011, Representative, Western Space and Facilities Committee  
2010-2011, Journal Club Coordinator, Western Social Psychology Area  
2009-2013, Graduate Campus Representative, Canadian Psychological Association  
2009-2010, Ph. D. Representative, Western Social Psychology Committee  
2008-2009, Journal Club Coordinator, Western Social Psychology Area  
2007-2008, M. Sc. Representative, Western Social Psychology Committee  
2007-2008, Lab Coordinator (sign-out system), Western Social Psychology Area  
2007-2008, Member, Western Psychology Graduate Students Association  
2006-2008, Member, Western Psychology Research and Technology Committee  
2006-2007, M. Sc. Representative, Western Undergraduate Affairs Committee  
2006-2007, Representative, Western Graduate Teaching Assistants' Union  
2004-2005, Volunteer Research Pool Coordinator, Simon Fraser Psychology Department